77.454

Bound
Periodical

Kansas City
Public Library

THE SCHOOL DISTRICT OF KANSAS CITY, MISSOURI.

This Volume is for

REFERENCE USE ONLY

5-32—6m—P

PUBLIC LIBRARY
KANSAS CITY
MO

From the collection of the

Prelinger Library

San Francisco, California
2008

VOLUME XI JANUARY, 1932 NUMBER 1

PUBLIC LIBRARY
KANSAS CITY
MO

THE BELL SYSTEM
TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS
OF ELECTRICAL COMMUNICATION

The Cathode Ray Oscillograph—*J. B. Johnson* . . . 1

The Operation of Vacuum Tubes as Class B and Class C Amplifiers—*C. E. Fay* 28

The Time Factor in Telephone Transmission—*O. B. Blackwell* 53

Constant Frequency Oscillators—*F. B. Llewellyn* . . 67

Some Physical Properties of Wiping Solders—*D. A. McLean, R. L. Peek, Jr., and E. E. Schumacher* 101

Regeneration Theory—*H. Nyquist* 126

Contemporary Advances in Physics, XXIII—Data and Nature of Cosmic Rays—*Karl K. Darrow* . . . 148

Abstracts of Technical Papers 185

Contributors to this Issue 195

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50c per Copy *$1.50 per Year*

PUBLIC LIBRARY
KANSAS CITY
MO

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the
American Telephone and Telegraph Company
195 Broadway, New York, N. Y.

EDITORIAL BOARD

Bancroft Gherardi F. B. Jewett
H. P. Charlesworth W. H. Harrison E. H. Colpitts
L. F. Morehouse H. D. Arnold O. B. Blackwell
Philander Norton, *Editor* J. O. Perrine, *Associate Editor*

SUBSCRIPTIONS

Subscriptions are accepted at $1.50 per year. Single copies are fifty cents each. The foreign postage is 35 cents per year or 9 cents per copy.

Copyright, 1932

PRINTED IN U. S. A.

THE BELL SYSTEM TECHNICAL JOURNAL

A JOURNAL DEVOTED TO THE
SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL
COMMUNICATION

EDITORIAL BOARD

Bancroft Gherardi	H. P. Charlesworth	F. B. Jewett
L. F. Morehouse	O. B. Blackwell	H. D. Arnold
W. H. Harrison	D. Levinger	H. S. Osborne

Philander Norton, *Editor* J. O. Perrine, *Associate Editor*

TABLE OF CONTENTS
AND
INDEX

VOLUME XI

1932

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

PRINTED IN U. S. A.

Bound
Periodical

Mr 9 '32

770454

THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME XI, 1932

Table of Contents

JANUARY, 1932

The Cathode Ray Oscillograph—*J. B. Johnson* 1
The Operation of Vacuum Tubes as Class B and Class C Amplifiers—*C. E. Fay* 28
The Time Factor in Telephone Transmission—*O. B. Blackwell* ... 53
Constant Frequency Oscillators—*F. B. Llewellyn* 67
Some Physical Properties of Wiping Solders—*D. A. McLean, R. L. Peck, Jr., and E. E. Schumacher* 101
Regeneration Theory—*H. Nyquist* 126
Contemporary Advances in Physics, XXIII—Data and Nature of Cosmic Rays—*Karl K. Darrow* 148

APRIL, 1932

A New Key West–Havana Carrier Telephone Cable—*H. A. Affel, W. S. Gorton and R. W. Chesnut* 197
Cellulose Acetate Treatment for Textile Insulation—Engineering Development—*E. B. Wood and D. R. Brobst* 213
Cellulose Acetate Treatment for Textile Insulation—Development of the Manufacturing Process—*C. R. Avery and H. Kress* ... 231
The Development of a Handset for Telephone Stations—*W. C. Jones and A. H. Inglis* 245
Precision Methods Used in Constructing Electric Wave Filters for Carrier Systems—*G. R. Harris* 264
Cathode Sputtering—A Commercial Application—*Hal F. Fruth* .. 283
A Voice and Ear for Telephone Measurements—*A. H. Inglis, C. H. G. Gray and R. T. Jenkins* 293

JULY, 1932

Henry as an Electrical Pioneer—*Bancroft Gherardi* 327
The Caesium-Oxygen-Silver Photoelectric Cell—*C. H. Prescott, Jr., and M. J. Kelly* 334
Two-Way Radio Telephone Circuits—*S. B. Wright and D. Mitchell* 368
Magnet Steels and Permanent Magnets—Relationships Among Their Magnetic Properties—*K. L. Scott* 383
A Method of Measuring Acoustic Impedance—*P. B. Flanders* ... 402
Transmission Lines for Short-Wave Radio Systems—*E. J. Sterba and C. B. Feldman* 411
An Efficient Miniature Condenser Microphone System—*H. C. Harrison and P. B. Flanders* 451
Wire Communication Aids to Air Transportation—*H. H. Nance* 462

OCTOBER, 1932

World-Wide Telephony—Its Problems and Future—*B. Gherardi and F. B. Jewett* 485
Long Distance Telephone Circuits in Cable—*A. B. Clark and H. S. Osborne* 520
The Conception and Demonstration of Electron Waves—*C. J. Davisson* 546
An Expansion for Laplacian Integrals in Terms of Incomplete Gamma Functions, and Some Applications—*E. C. Molina* ... 563
Contemporary Advances in Physics, XXIV—High-Frequency Phenomena in Gases, First Part—*Karl K. Darrow* 576
Transformer Coupling Circuits for High-Frequency Amplifiers—*A. J. Christopher* 608

Index to Volume XI

A

Acoustic Impedance, A Method of Measuring, *P. B. Flanders,* page 402.

Affel, H. A., W. S. Gorton and R. W. Chesnut, A New Key West–Havana Carrier Telephone Cable, page 197.

Air Transportation, Wire Communication Aids to, *H. H. Nance,* page 462.

Amplifiers, Class B and Class C, The Operation of Vacuum Tubes as, *C. E. Fay,* page 28.

Amplifiers, High-Frequency, Transformer Coupling Circuits for, *A. J. Christopher,* page 608.

Avery, C. R. and H. Kress, Cellulose Acetate Treatment for Textile Insulation—Development of the Manufacturing Process, page 231.

B

Blackwell, O. B., The Time Factor in Telephone Transmission, page 53.

Brobst, D. R. and E. B. Wood, Cellulose Acetate Treatment for Textile Insulation—Engineering Development, page 213.

C

Cable, Long Distance Telephone Circuits in, *A. B. Clark and H. S. Osborne,* page 520.

Cable, A New Key West–Havana Carrier Telephone, *H. A. Affel, W. S. Gorton and R. W. Chesnut,* page 197.

Cæsium-Oxygen-Silver Photoelectric Cell, The, *C. H. Prescott, Jr. and M. J. Kelly,* page 334.

Carrier Systems, Precision Methods Used in Constructing Electric Wave Filters for, *G. R. Harris,* page 264.

Carrier Telephone Cable, A New Key West–Havana, *H. A. Affel, W. S. Gorton and R. W. Chesnut,* page 197.

Cathode Ray Oscillograph, The, *J. B. Johnson,* page 1.

Cathode Sputtering—A Commercial Application, *Hal F. Fruth,* page 283.

Cellulose Acetate Treatment for Textile Insulation—Engineering Development, *E. B. Wood and D. R. Brobst,* page 213.

Cellulose Acetate Treatment for Textile Insulation—Development of the Manufacturing Process, *C. R. Avery and H. Kress,* page 231.

Chesnut, R. W., W. S. Gorton and H. A. Affel, A New Key West–Havana Carrier Telephone Cable, page 197.

Christopher, A. J., Transformer Coupling Circuits for High-Frequency Amplifiers, page 608.

Clark, A. B. and H. S. Osborne, Long Distance Telephone Circuits in Cable, page 520.

Conception and Demonstration of Electron Waves, The, *C. J. Davisson,* page 546.

Condenser Microphone System, An Efficient Miniature, *H. C. Harrison and P. B. Flanders,* page 451.

Constant Frequency Oscillators, *F. B. Llewellyn,* page 67.

Contemporary Advances in Physics, XXIII—Data and Nature of Cosmic Rays, *Karl K. Darrow,* page 148.

Contemporary Advances in Physics, XXIV—High-Frequency Phenomena in Gases, First Part, *Karl K. Darrow,* page 576.

D

Darrow, Karl K., Contemporary Advances in Physics, XXIII—Data and Nature of Cosmic Rays, page 148.
Contemporary Advances in Physics, XXIV—High-Frequency Phenomena in Gases, First Part, page 576.
Davisson, C. J., The Conception and Demonstration of Electron Waves, page 546.
Development of a Handset for Telephone Stations, The, *W. C. Jones and A. H. Inglis,* page 245.

E

Ear for Telephone Measurements, A Voice and, *A. H. Inglis, C. H. G. Gray and R. T. Jenkins,* page 293.
Efficient Miniature Condenser Microphone System, An, *H. C. Harrison and P. B. Flanders,* page 451.
Electron Waves, The Conception and Demonstration of, *C. J. Davisson,* page 546.
Expansion, An, for Laplacian Integrals in Terms of Incomplete Gamma Functions, and Some Applications, *E. C. Molina,* page 563.

F

Fay, C. E., The Operation of Vacuum Tubes as Class B and Class C Amplifiers, page 28.
Feldman, C. B. and E. J. Sterba, Transmission Lines for Short-Wave Radio Systems, page 411.
Filters, Electric Wave, for Carrier Systems, Precision Methods Used in Constructing, *G. R. Harris,* page 264.
Flanders, P. B., A Method of Measuring Acoustic Impedance, page 402.
Flanders, P. B. and H. C. Harrison, An Efficient Miniature Condenser Microphone System, page 451.
Fruth, Hal F., Cathode Sputtering—A Commercial Application, page 283.

G

Gherardi, Bancroft, Henry as an Electrical Pioneer, page 327.
Gherardi, Bancroft and Frank B. Jewett, World-Wide Telephony—Its Problems and Future, page 485.
Gorton, W. S., R. W. Chesnut and H. A. Affel, A New Key West–Havana Carrier Telephone Cable, page 197.
Gray, C. H. G., R. T. Jenkins and A. H. Inglis, A Voice and Ear for Telephone Measurements, page 293.

H

Handset for Telephone Stations, The Development of a, *W. C. Jones and A. H. Inglis,* page 245.
Harris, G. R., Precision Methods Used in Constructing Electric Wave Filters for Carrier Systems, page 264.
Harrison, H. C. and P. B. Flanders, An Efficient Miniature Condenser Microphone System, page 451.
Henry as an Electrical Pioneer, *Bancroft Gherardi,* page 327.

I

Inglis, A. H. and W. C. Jones, The Development of a Handset for Telephone Stations, page 245.
Inglis, A. H.. C. H. G. Gray and R. T. Jenkins, A Voice and Ear for Telephone Measurements, page 293.

Insulation, Textile, Cellulose Acetate Treatment for—Engineering Development, *E. B. Wood and D. R. Brobst,* page 213.

Insulation, Textile, Cellulose Acetate Treatment for—Development of the Manufacturing Process, *C. R. Avery and H. Kress,* page 231.

J

Jenkins, R. T., C. H. G. Gray and A. H. Inglis, A Voice and Ear for Telephone Measurements, page 293.

Jewett, Frank B. and Bancroft Gherardi, World-Wide Telephony—Its Problems and Future, page 485.

Johnson, J. B., The Cathode Ray Oscillograph, page 1.

Jones, W. C. and A. H. Inglis, The Development of a Handset for Telephone Stations, page 245.

K

Key West–Havana Carrier Telephone Cable, A New, *H. A. Affel, W. S. Gorton and R. W. Chesnut,* page 197.

Kelly, M. J. and C. H. Prescott, Jr., The Cæsium-Oxygen-Silver Photoelectric Cell, page 334.

Kress, H. and C. R. Avery, Cellulose Acetate Treatment for Textile Insulation—Development of the Manufacturing Process, page 231.

L

Laplacian Integrals in Terms of Incomplete Gamma Functions, An Expansion for, and Some Applications, *E. C. Molina,* page 563.

Llewellyn, F. B., Constant Frequency Oscillators, page 67.

Long Distance Telephone Circuits in Cable, *A. B. Clark and H. S. Osborne,* page 520.

M

McLean, D. A., R. L. Peek, Jr. and E. E. Schumacher, Some Physical Properties of Wiping Solders, page 101.

Magnet Steels and Permanent Magnets—Relationships Among Their Magnetic Properties, *K. L. Scott,* page 383.

Microphone System, An Efficient Miniature Condenser, *H. C. Harrison and P. B. Flanders,* page 451.

Mitchell, D. and S. B. Wright, Two-Way Radio Telephone Circuits, page 368.

Molina, E. C., An Expansion for Laplacian Integrals in Terms of Incomplete Gamma Functions, and Some Applications, page 563.

N

Nance, H. H., Wire Communication Aids to Air Transportation, page 462.

Nyquist, H., Regeneration Theory, page 126.

O

Osborne, H. S. and A. B. Clark, Long Distance Telephone Circuits in Cable, page 520.

Oscillators, Constant Frequency, *F. B. Llewellyn,* page 67.

Oscillograph, The Cathode Ray, *J. B. Johnson,* page 1.

P

Peek, R. L., Jr., E. E. Schumacher and D. A. McLean, Some Physical Properties of Wiping Solders, page 101.

Photoelectric Cell, The Cæsium-Oxygen-Silver, *C. H. Prescott, Jr., and M. J. Kelly,* page 334.

Physical Properties of Wiping Solders, Some, *D. A. McLean, R. L. Peek, Jr., and E. E. Schumacher,* page 101.

Physics, XXIII, Contemporary Advances in—Data and Nature of Cosmic Rays, *Karl K. Darrow,* page 148.

Physics, XXIV, Contemporary Advances in—High-Frequency Phenomena in Gases, First Part, *Karl K. Darrow,* page 576.

Prescott, C. H., Jr., and M. J. Kelly, The Cæsium-Oxygen-Silver Photoelectric Cell, page 334.

R

Radio Systems, Short-Wave, Transmission Lines for, *E. J. Sterba and C. B. Feldman,* page 411.

Radio Telephone Circuits, Two-Way, *S. B. Wright and D. Mitchell,* page 368.

Regeneration Theory, *H. Nyquist,* page 126.

S

Schumacher, E. E., R. L. Peek, Jr., and D. A. McLean, Some Physical Properties of Wiping Solders, page 101.

Scott, K. L., Magnet Steels and Permanent Magnets—Relationships Among Their Magnetic Properties, page 383.

Solders, Wiping, Some Physical Properties of, *D. A. McLean, R. L. Peek, Jr., and E. E. Schumacher,* page 101.

Sterba, E. J. and C. B. Feldman, Transmission Lines for Short-Wave Radio Systems, page 411.

T

Time Factor in Telephone Transmission, The, *O. B. Blackwell,* page 53.

Transformer Coupling Circuits for High-Frequency Amplifiers, *A. J. Christopher,* page 608.

Transmission Lines for Short-Wave Radio Systems, *E. J. Sterba and C. B. Feldman,* page 411.

V

Vacuum Tubes as Class B and Class C Amplifiers, The Operation of, *C. E. Fay,* page 28.

Voice and Ear for Telephone Measurements, A, *A. H. Inglis, C. H. G. Gray and R. T. Jenkins,* page 293.

W

Wire Communication Aids to Air Transportation, *H. H. Nance,* page 462.

Wood, E. B. and D. R. Brobst, Cellulose Acetate Treatment for Textile Insulation—Engineering Development, page 213.

World-Wide Telephony—Its Problems and Future, *Bancroft Gherardi and Frank B. Jewett,* page 485.

Wright, S. B. and D. Mitchell, Two-Way Radio Telephone Circuits, page 368.

The Bell System Technical Journal

January, 1932

The Cathode Ray Oscillograph *

By J. B. JOHNSON

The cathode ray oscillograph, since its invention by Braun, has developed along three lines. The major types of tubes are the high voltage tubes with a fluorescent screen, the high voltage tubes with internal photographic equipment, and the low voltage tubes. This paper follows the structural development of commercial tubes. The operation of the tubes is discussed, from the standpoint of both theory and practice, with particular reference to the low voltage type of tube. Numerous examples are given of the applications of the tubes to problems in science and engineering.

IN our complicated life, we find that we need a great many aids to our primary sense organs. The processes of the modern world demand that we make correct estimates of things that are too large or too small, too intense or too feeble, for our poor senses. We have balances to give us the weight of masses too heavy for us to lift or too small to be felt. Telescopes enable us to see far-off objects, microscopes very small objects. Our ears are supplemented by telephones that put us within earshot of almost all the civilized world. For electric currents we have ammeters to measure currents so large as to destroy us in a second, and galvanometers that measure currents far too small for us to feel as a shock. Taste and smell have not yet been supplied with artificial aids, but that may come some day.

For recording long times we have clocks and calendars; for making a record of happenings that take place in a time too short for us to think of, we use oscillographs.

There are a number of different types of oscillographs in use, all of them electrical in nature. The kind I am going to discuss involves a stream of cathode rays and it is therefore called the cathode ray oscillograph. The principle of its operation is quite simple. We have two electrodes in an elongated, evacuated glass tube as in Fig. 1; one of them may be a heated filament, the other a plate with a small hole in it. When a potential is applied between the electrodes, making the filament cathode and the plate anode, the electrons emitted by the hot filament are drawn to the anode. Some of them pass through the fine hole in the anode and continue as a thin pencil of electrons, a cathode ray, down the length of the tube. At the end of the tube is a screen of fluorescent materials, which shines brightly at the point where the ray strikes it. We can therefore see where the ray ends on the screen. Another pair of electrodes in the form of two plates P and P_1 is introduced so that the cathode ray passes between them (Fig. 2). Now, by a battery or otherwise, we apply a voltage between the plates, so that one is positive with respect to the other. The electrons of the ray, being negative charges, are during their passage between the plates drawn toward the positive plate and emerge in a different direction because of the applied voltage. Similarly, a magnetic field applied by the magnet N-S, across the path of the ray and in the plane of the paper, as in Fig. 3, makes the ray emerge in a direction out from the page. The amount of the deflection is a measure of the strength of the applied magnetic or electric field. We have, then, in this cathode ray a pointer which tells the magnitude of the field that deflects it. It is, furthermore, a pointer that is almost without mass and sluggishness; it is almost not a material pointer. It can therefore follow variations in the applied field that are very rapid, as we shall see presently. Because of this property the instrument has been used extensively for studying the electrical phenomena of such a range of subjects as power machinery, telephone apparatus, radio transmission and electric surges.

---

* Presented at Franklin Inst. mtg., Dec. 4, 1930. *Jour. Franklin Inst.*, December, 1931.

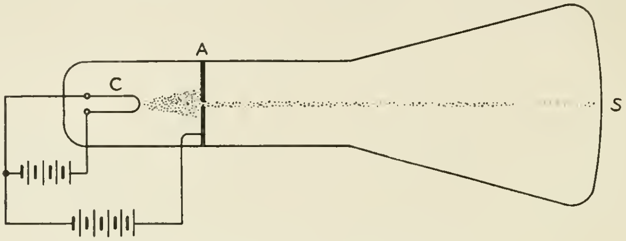


Fig. 1—Schematic of cathode ray tube.

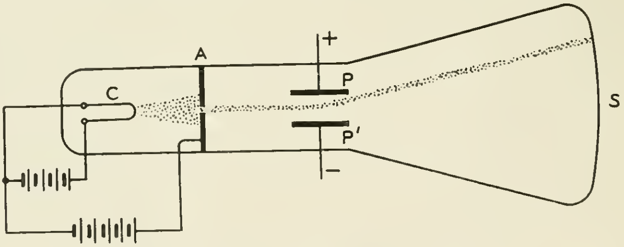


Fig. 2—Tube with electrostatic deflection.

One of the more dramatic recent applications has been the investigation of lightning, probably the most important work on lightning since its electrical nature was discovered by Benjamin Franklin one hundred eighty years ago.

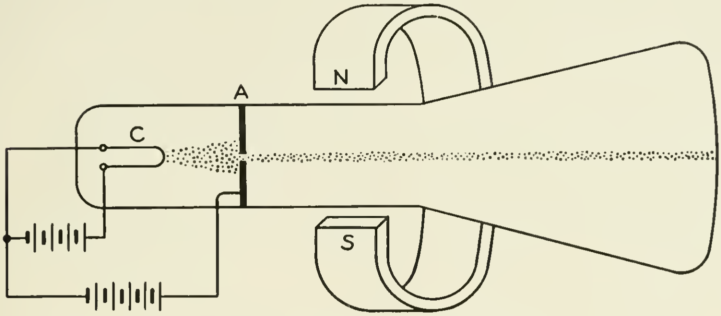


Fig. 3—Tube with magnetic deflection.

Let us examine the operation of the tube more closely. The speed of the electrons as they emerge from the aperture in the anode can be determined from the energy equation,

$$\tfrac{1}{2}mv^2 = eV,$$

the energy of motion equaling the total work done on the electron by the field between cathode and anode. V is the potential between cathode and anode, e the electric charge constituting the electron, m its mass and v its speed. Solving for the speed we have the relation between the speed and the driving voltage,

$$v = \sqrt{2\frac{e}{m}V}.$$

The value of e/m is known to be 1.77×10^7 e.m.u., the volt is 10^8 e.m.u. and the velocity of the electrons is thus

$$v = 5.95 \times 10^7\sqrt{V}\text{ cm./sec.}$$

If the driving potential is 300 volts, then the speed of the electrons is given as roughly 1×10^9 cm./sec. or 6000 miles per second. For a tube 20 cm. long an electron travels from the deflector plates to the screen in $20/10^9 = 1/50{,}000{,}000$ sec. If the applied voltage is 30,000 volts the speed is very nearly ten times as great as with 300 volts; it is ⅓ the velocity of light. A change of direction of the ray induced at the deflector plates is therefore transmitted to the end of the ray in a very short time, and the ray can follow faithfully potential variations at the plates that are very rapid.

Let us see how the ray responds to voltage applied to the plates. The ray normally travels with a speed v along the tube. Referring to Fig. 4, the ray now passes between two plates of length l and separation d, between which a potential difference V' is maintained. While the ray is passing between the plates the electrons are subject to an acceleration,

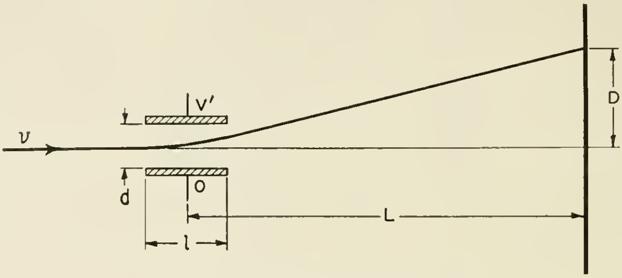


Fig. 4—Electrostatic deflection.

$$a = \frac{e}{m}E = \frac{e}{m}\frac{V'}{d}.$$

This continues during the time $t = l/v$. The transverse velocity acquired is therefore

$$v' = at = \frac{e}{m}\frac{V'}{d}\frac{l}{v}.$$

The ray then travels on in a straight line to the screen which it meets at a distance D from the normal position. The deflection D bears the same relation to the length of the beam, from the center of the deflecting plates, as the transverse velocity bears to the longitudinal.

$$\frac{D}{L} = \frac{v'}{v} = \frac{e}{m}\frac{V'}{d}\frac{l}{2\frac{e}{m}V} = \frac{1}{2}\frac{l}{d}\frac{V'}{V},$$

$$D = \frac{1}{2}\frac{lL}{d}\frac{V'}{V}.$$

This brings out an interesting point in connection with the design of the deflector plates. For high sensitivity the plates should be long and close together but the plates must not cut the path of the deflected ray. If we want to get a certain maximum deflection D with a tube of a certain length L, then the relation of spacing to length of the plates must be

$$\frac{d}{l} = \frac{D}{L},$$

as can easily be seen from Fig. 4. When this condition is satisfied it makes little difference in the sensitivity whether the plates are large and far apart or small and close together.

The magnetic sensitivity of the tube is more uncertain only because the boundaries of the magnetic field are usually less well defined than those of the electric field. One form of the derivation will be given here. If the ray generated by the driving voltage V passes for a distance l (Fig. 5) through a region in which there is a transverse magnetic field of strength H, the path of the beam is an arc of a circle whose radius is

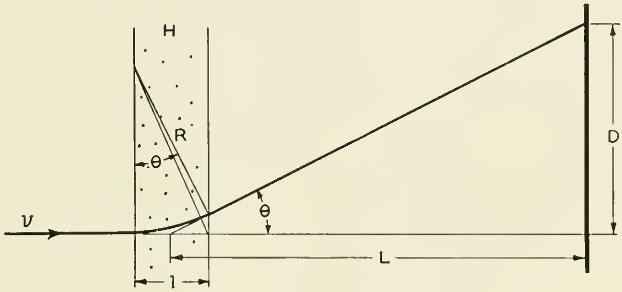


Fig. 5—Magnetic deflection.

$$R = \frac{mv}{eH} = \frac{1}{H}\sqrt{\frac{2mV}{e}}\,.$$

When the ray leaves the magnetic region it proceeds in a straight line to the screen, where the total deflection from the normal is D. If the angular deflection θ is not great then we have, very nearly,

$$\tan\theta = \frac{D}{L} \doteq \frac{l}{R} = lH\sqrt{\frac{e}{2mV}}$$

and hence

$$D \doteq lLH\sqrt{\frac{e}{2mV}}\,.$$

In practical units instead of electromagnetic the expression becomes

$$D \doteq .3lLH/\sqrt{V}.$$

Having now described in an elementary way how the cathode ray oscillograph works, let me turn to the story of the development of the tube.[1] The first reference I have seen to the idea that a cathode ray might be used to indicate magnetic field dates to 1894, when Hess,[2] in France, suggested the use of such a tube as a curve tracer. The first application of the idea, however, was made by Ferdinand Braun [3] in 1897, and after him the instruments have been called Braun tubes.

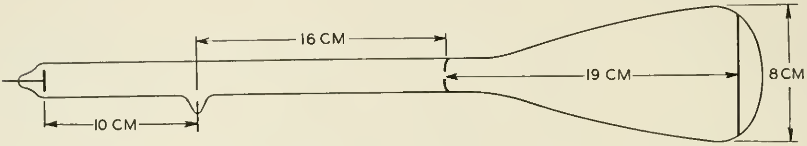


Fig. 6—The first cathode ray oscillograph, F. Braun, 1897.

The tube of Braun was quite simple (Fig. 6). It had a flat disc cathode, a wire in a side tube as anode, a pierced diaphragm to limit the beam and a fluorescent screen of zinc sulphide. It contained air at low pressure. Current from an electrostatic machine produced a discharge in the residual gas in the tube, from which emanated the cathode rays through the aperture in the diaphragm.

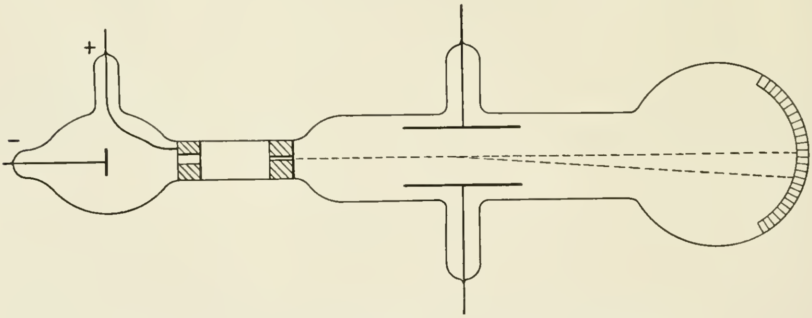


Fig. 7—Tube for measuring e/m, J. J. Thomson, 1897.

It is of interest that the invention of the tube took place before the nature of cathode rays was understood. It was in the same year that J. J. Thomson in England and W. Kauffmann in Germany, each using a tube that was almost identical with the Braun tube, determined that cathode rays have mass as well as charge. Thomson's tube [4] is shown in Fig. 7.

---

[1] More detailed chronicles of the development of the tube have been made by H. Hausrath, "Apparatus and Technique for Producing and Recording Curves of Alternating Currents and Electrical Oscillations," *Helios*, 1912; and by MacGregor-Morris and Mines, "Measurements in Electrical Engineering by Means of Cathode Rays," *Jl. Inst. El. Eng.*, **63**, p. 1056, 1925.

[2] Hess, A., *Compt. Rend.*, **119**, p. 57, 1894.

[3] Braun, Ferdinand, *Wied. Ann.*, **60**, p. 552, 1897.

The Braun tube immediately found many applications. One of the most fruitful fields for it was its use by Professor Zenneck in studying radio circuits and the transmission of radio waves. From Professor Zenneck and his school there are still papers coming out on work done by means of the Braun tube. The tube was introduced in this country very early. Professor H. J. Ryan, then at Cornell, in 1903 described measurements on high voltage power circuits, and similar work has appeared occasionally ever since from Professor Ryan's hand.

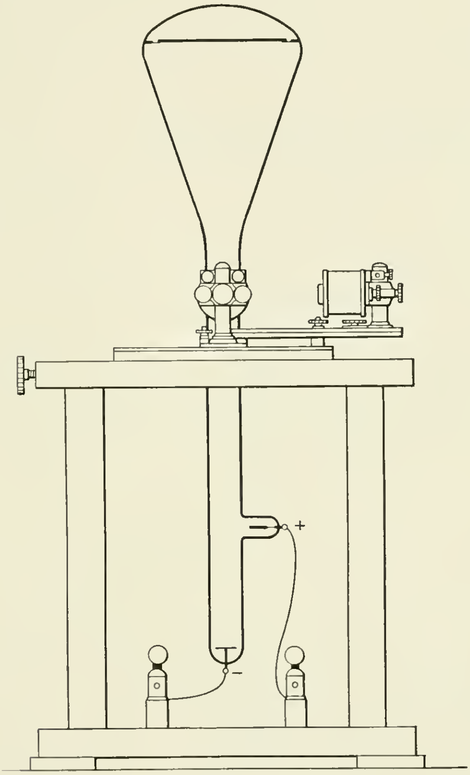


Fig. 8—Ebert and Hoffmann, 1898—tube by Geissler.[5]

After the invention of the tube there were a number of improvements that made it more convenient and its operation more reliable. Figures 8–12 show some of the many designs of tubes of this time.

---

[4] Thomson, J. J., *Phil. Mag.*, **44**, p. 293, 1897.
[5] Ebert and Hoffmann, *E. T. Z.*, **19**, p. 405, 1898.

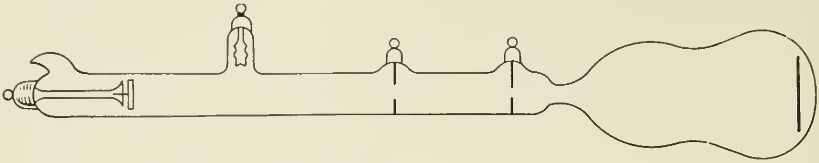


Fig. 9—MacGregor-Morris, 1902—tube by Cossor.[6]

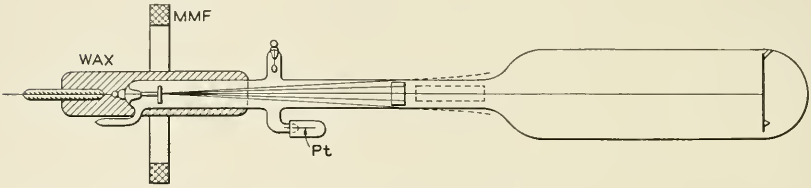


Fig. 10—Ryan, 1903—tube by Muller-Uri.[7]

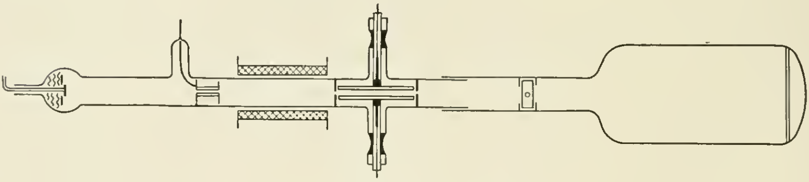


Fig. 11—Roschansky, 1911.[8]

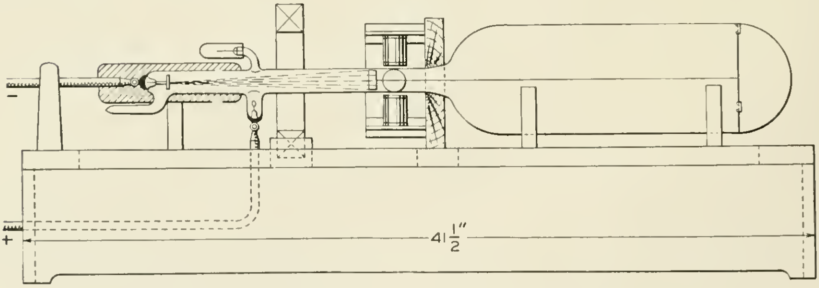


Fig. 12—Broughton, 1913—tube by Max Kohl.[9]

---

[6] MacGregor-Morris, *Engineering*, 1902, **73**, p. 754.
[7] Ryan, H. J., *Am. Inst. El. Eng., Trans.*, **22**, p. 539, 1903.
[8] Roschansky, D., *Ann. d. Phys.*, **36**, p. 281, 1911.
[9] Broughton, H. H., *Electrician*, **72**, p. 171, 1913.

In 1905 Wehnelt[10] suggested the use of a hot, lime coated filament, which he had found a couple of years earlier to be a strong emitter of electrons and which was the basis of the present day oxide coated filaments in vacuum tubes. Wehnelt made up such a tube (Fig. 13) which he could operate on the 220 volt power circuit. This was probably the first practical application of the oxide coated filament. A number of experimental tubes were made up with hot filaments in the years following, but almost twenty years were to elapse before a really successful tube with hot cathode and low voltage was developed. This was the Western Electric 224 tube,[11] to be discussed in greater detail presently.

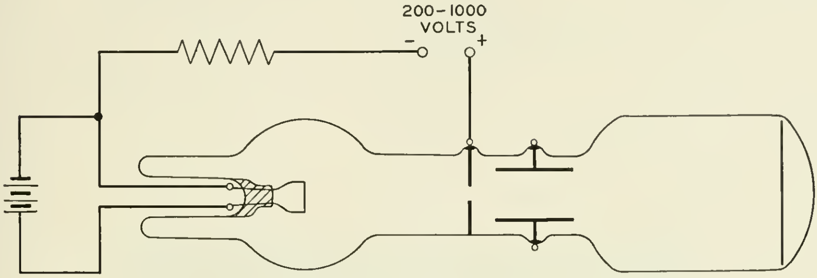


Fig. 13—Wehnelt, 1905.

Another tube of this type, by von Ardenne and Hartel,[12] is illustrated in Fig. 14.

After Braun and Wehnelt, the most notable change in structure was made by A. Dufour[13] in France in 1914. Up to that time permanent records of the pattern on the fluorescent screen were made by means of a camera. This usually meant that the pattern for a rapid phenomenon had to be repeated many times before the photographic plate was exposed enough. Dufour omitted the fluorescent screen and instead placed the photographic plate inside the tube so that the cathode ray could play directly on it. When the cathode rays strike the photographic emulsion directly a record can be traced in a much shorter time than when the intermediary light from a fluorescent screen is focused by a lens on the photographic plate. Placing the photographic plates internally of course involved a number of complications, such as mechanism for moving the plates inside the evacuated tube, means for inserting and taking out the plates, and pumps for producing and maintaining the vacuum. The old glass structure

[10] Wehnelt, A., *Phys. Zeit.*, **6**, p. 732, 1905.
[11] Johnson, J. B., *Jl. Am. Opt. Soc. & R. S. I.*, **6**, p. 701, 1922.
[12] Hartel, H. von, *Zeit. f. Hochfr. Techn.*, **34**, p. 227, 1929.
[13] Dufour, A., *Compte Rend.*, **158**, p. 1339, 1914.

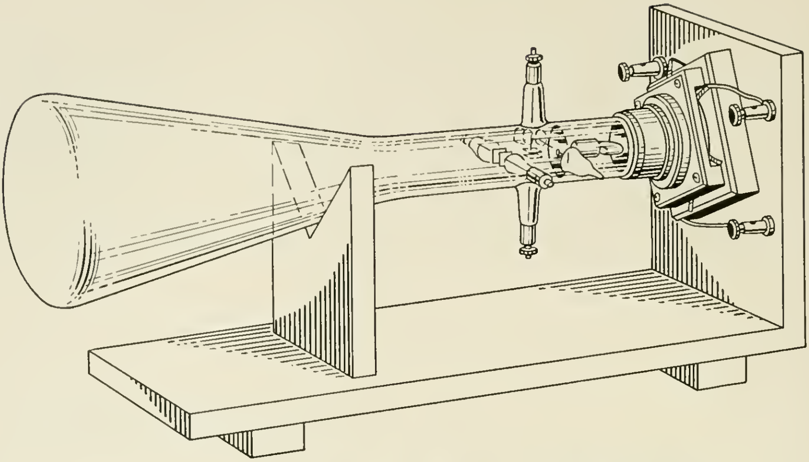


Fig. 14—von Ardenne-Hartel, 1930—tube by Leybold's Nachfolger.

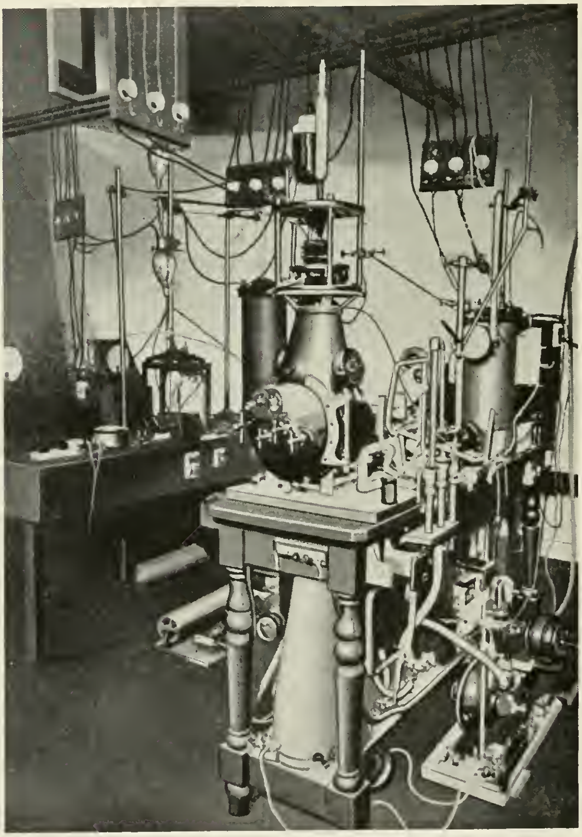


Fig. 15—Dufour oscillograph of 1923.[14]

---

[14] Dufour, A., Oscillographe Cathodique, Etienne Chiron, Paris, 1923.

is largely abandoned in these tubes, and metal is substituted. The applied voltage is very high, of the order of 50,000 volts. This makes a rather formidable piece of apparatus (Fig. 15), but quite a useful one.

In the last few years several different tubes of this type have been developed. Aside from Dufour's tube, there are those of Rogowski in Germany, Wood in England, Berger in Switzerland, Norinder in Sweden, The Westinghouse Co. and the General Electric Co. in this country (Figs. 16–18, 21). All of them involve complicated tubes and control circuits. Some tubes are made to operate during a single peak of a 60 cycle wave. With others there are ingenious switching devices that start the tube operating at the very beginning of the electrical impulse to be studied, and the tube then proceeds to record the rest of the impulse. When we consider that the impulse may be a stroke of lightning on a transmission line, we realize that there are some things that are "faster than greased lightning."

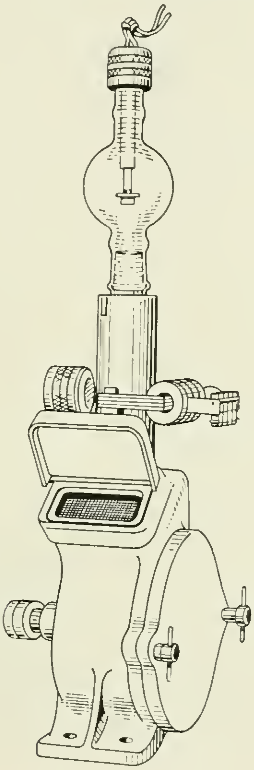


Fig. 16—Wood, 1923.[15]

---

[15] Wood, A. B., *Phys. Soc. Lond. Proc.*, **35**–**2**, p. 109, 1923.

In the last couple of years a further step has been made by Max Knoll.[16] He seems to have simplified the operation of the tube considerably by attaching on the end of his tube a thin window in the manner of Dr. Coolidge's cathode ray tube. In this way the photographic plate on the outside of the tube is exposed to the pencil of cathode rays that has come through the thin window of metal or of cellophane.

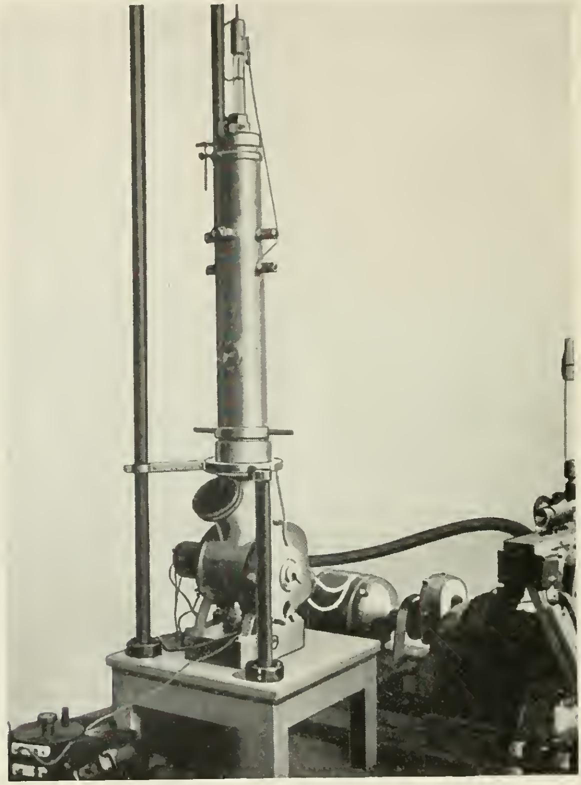


Fig. 17—Westinghouse-Norinder, 1928.[17]

We have, then, three general classes of cathode ray oscillographs: First those that resemble the original Braun tube, still in limited use, comprising a glass tube with a fluorescent screen, and a relatively high operating voltage; second the tubes of the Dufour type that I have just described, with direct recording on the photographic plate or film; and third, tubes with a hot cathode, with a relatively low operating voltage.

---

[16] Knoll, Max, *Zeits. f. tech. Phys.*, **10**, p. 28, 1929.
[17] Norinder, H., *A. I. E. E. Trans.*, **47**, p. 446, 1928.

Of this last type is the Western Electric No. 224 Cathode Ray Oscillograph. Since this is the tube with which I have had more direct contact I wish to discuss some problems of its development and operation.

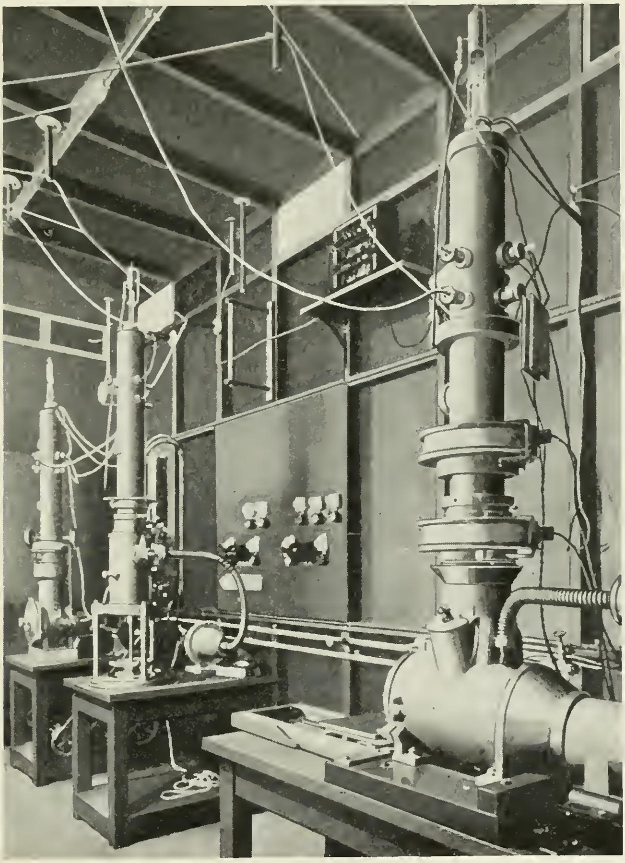


Fig. 18—Norinder, 1930—group of oscillographs.[18]

The design of the tube shown in Fig. 19 is now fairly familiar,* and

---

[18] Norinder, H., *Zeits. f. Phys.*, **63**, p. 672, 1930.

* Since the paper was read the tube has been altered in certain respects. The glass enclosure for the filament has been replaced by one of metal, and the anode and deflector plates are supported on machined insulating blocks. These changes make the structure more rugged and insure more nearly perfect alignment of the parts. The end of the bulb has been changed so as to present a cylindrical surface instead of a spherical one, thus permitting more intimate contact between the fluorescent screen and a photographic film when contact photographs are made.

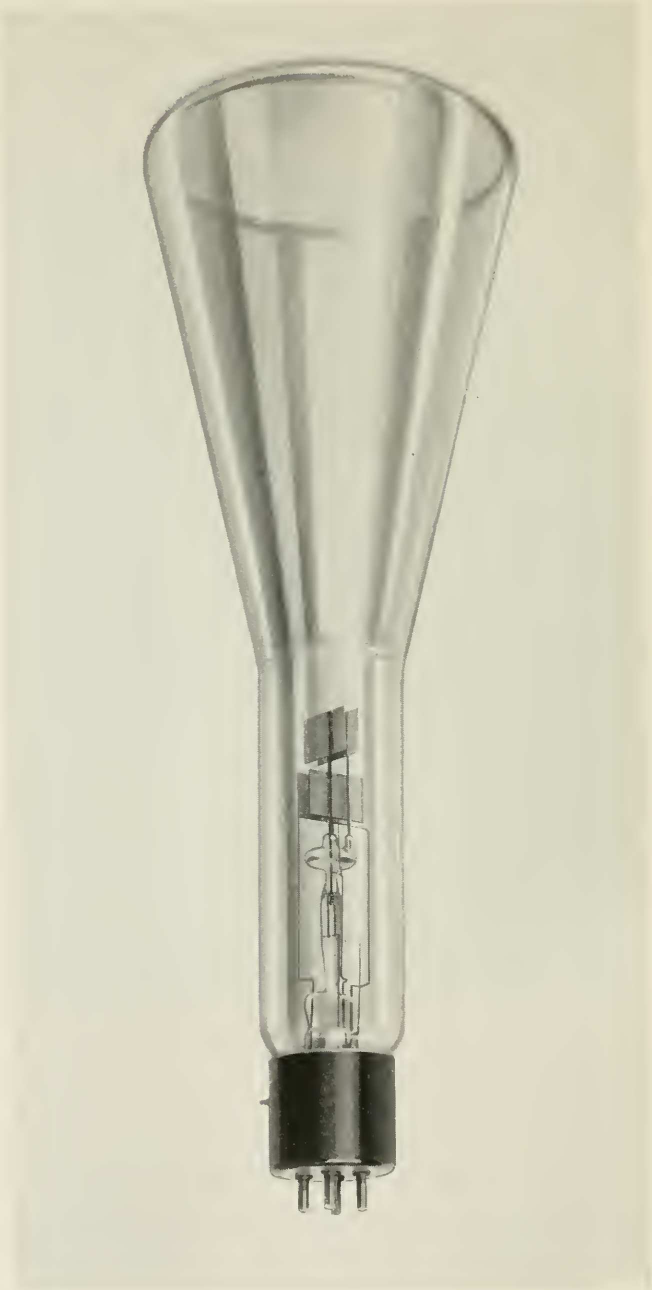


Fig. 19—Western Electric tube.

I shall describe it only as I describe the reason for the various features.

First we wanted a convenient tube to operate on the batteries of an ordinary vacuum tube. The thermionic filament cathode was therefore required. The anode is a metal tube placed a short distance from the cathode and between them is a metal disc with a perforation through which the electrons pass to the anode. These electrodes are represented respectively by the letters *C*, *A* and *D* in Fig. 20.

The electrons flow from the cathode to the inside of the anode and some of them pass through and form the electron beam. Now, for reasons to be mentioned presently, there is some gas in the tube and this puts two requirements on the structure of the electron gun. First, the gas would permit considerable ionization in the tube if the electrons were permitted to pass into it. Most of the current would go around the disc and to the outside of the anode. The cathode and anode are therefore enclosed in a small tube, having dimensions less than the mean free path of electrons in the gas so that no appreciable ionization can build up. There is some ionization, however, in the space between the cathode and anode, and the positive ions produced tend to bombard the filament. If the filament is directly exposed to this bombardment the oxide coating is worn off, as in a sand blast, in a matter of two or three hours. The filament is therefore wound in the shape of a helix which is mounted coaxially with the anode and the perforation in the disc so as to be out of the direct path of the ions. In this way the filament is made to last several hundred hours of operation.

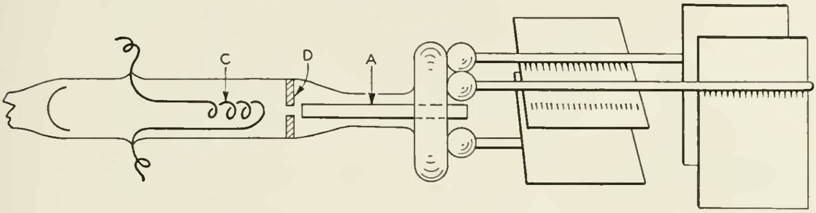


Fig. 20—Diagram of electron gun.

This completes the internal parts of the electron gun. Externally it has mounted on it two pairs of deflector plates that control the direction of the electrons after they leave the gun. In order to prevent any large difference of potential between the anode and deflector plates, one plate of each pair is connected directly to the anode, and only the other plate has the variable potential impressed on it. As for size and separation, the plates are designed to give maximum sensitivity for a given full deflection. The sensitivity is about 1 mm.

deflection for each volt applied to the deflector plates, or 1 mm. for each ampere turn in a pair of small coils placed outside of the tube. These figures are for the normal driving potential of 300 volts.

The flattened end of the bulb, on which the electron beam impinges, is covered with a fluorescent material. The powder is a mixture of zinc orthosilicate and calcium tungstate, both specially prepared for fluorescence. The zinc silicate produces a green light of high visibility and the tungstate a blue light of high photographic activity, so that with the mixture the same tubes can be used efficiently for both visual and photographic observations.

As said before, there is some gas in the tube. One purpose of the gas is to produce a small amount of ionization in the tube which prevents any unduly large charges from accumulating on the glass walls and screen. Electrons deposited on the glass are neutralized by positive ions produced in the gas. An electron current equal to the current in the beam drifts back through the gas to the anode. The chief result of this drift of electrons is that a negative space charge is formed in the tube which decreases the speed of the electrons before they strike the screen as if the driving potential had been lowered by about 50 volts.

The other and more important purpose of the gas is to bring the electrons of the beam to a sharper focus at the screen. The beam is diffuse for two reasons, first because it is originally divergent, and secondly because of the natural electrostatic repulsion that tends to force the electrons apart. The gas serves to overcome these actions in the following way: As the beam of electrons travels down the length of the tube, some electrons collide with atoms of the gas and separate the atom into an electron and a positive ion. The impact of the electron does little to displace the massive positive ion from the position it temporarily occupies while two electrons are immediately shot out of the path. The result is a column of positive ionization down the length of the beam, with a negative space charge surrounding it. This produces a radial electrostatic field which tends to bend the path of the outer electrons of the beam inward toward the centre. The magnitude of this action depends on the degree of the differential ionization. This, again, is the greater the higher the gas pressure and the greater the current in the beam. The gas pressure must be low enough so that the larger fraction of the electrons reach the screen, and then the current in the beam must be such as to produce the desired focusing action. The heavier the ions the lower can the pressure be. The condition for a focus then involves the kind and pressure of gas, the speed of the electrons, the current in the beam and the length of the tube. In the 224 tube, with argon at the pressure of .01 mm. the focusing occurs at about 20 microamperes in the beam. Less current makes a large unfocused spot; more current produces a focus before the screen is reached, with consequent spreading of the spot again. Hence it is that the spot on the fluorescent screen is focused by adjusting the heating current of the cathode.

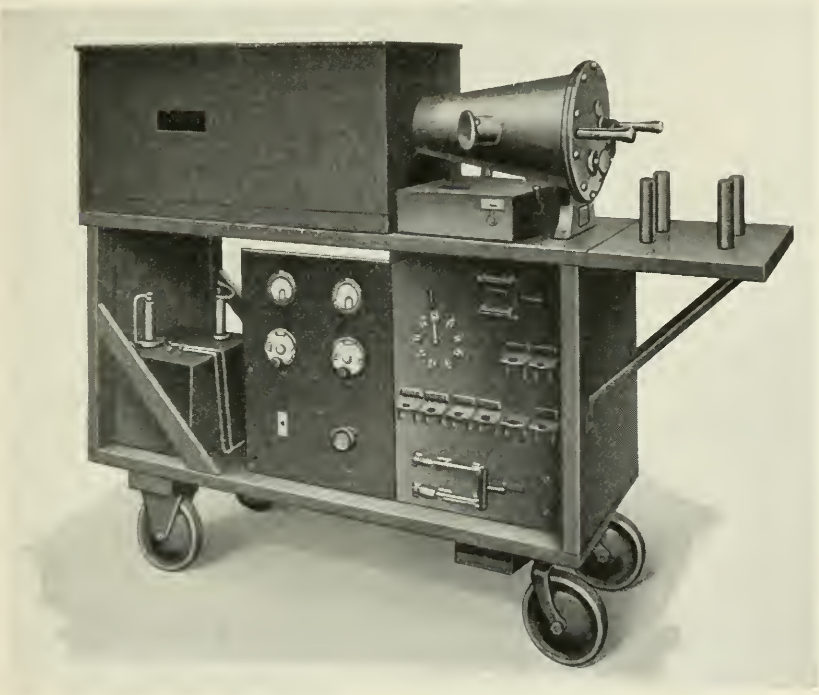


Fig. 21—General Electric, 1928.[19]

Besides focusing the electrons and preventing the accumulation of charges in the tube, the gas plays various other roles. One very curious effect of the gas is that it decreases the sensitivity of the tube at small deflecting voltages. When uniformly varying voltage is applied to a pair of deflector plates so as to move the spot across the screen, the spot seems to hesitate for an instant at the centre of the screen, appearing to be brighter there. The effect has always been observed but particular attention has recently been given to it by Professor Bedell.[20] The explanation has to do with space charge be-

[19] Lee, E. S., *G. E. Rev.*, **31**, p. 404, 1928.

[20] Bedell, F. & Kuhn, J., *Phys. Rev.* **36**, p. 993, 1930.

tween the deflector plates. The beam of electrons produces slowly moving positive ions and electrons in the gas along its path between the plates. When a voltage is applied to the plates, the positive ions travel from the beam toward the negative plate and an equal number of the electrons travel toward the positive plate. The space charge set up by the electrons and ions produces an electric field opposing that created by the applied voltage. The greatest space charge occurs near the negative plate where the sluggish positive ions flow. The space midway between the plates remains nearly field free and there is little deflection of the beam until the voltage is greater than that at which all of the ions produced are drawn to the plate. Calculation agrees with observation that this voltage is 2 to 3 volts on either side of zero.

Having now described the operation and structure of some of the oscillographs, I should like to say something more about their uses.

The cathode ray oscillograph is essentially a curve tracer that plots out in rectangular coordinates the relation between two quantities represented by the fields between the deflector plates. Often one of these quantities is time, as in the ordinary moving mirror oscillograph, while the other is some electrical quantity. We then say that we plot a wave shape. We must then have a way of making the spot move at a uniform rate. One of the simplest and most reliable ways of producing a linear time axis,[21] at least for low voltage tubes, employs two thermionic tubes to control the charge in a condenser as shown in Fig. 22. One tube, a simple two-electrode tube T_1, limits the charging current of the condenser C so that the voltage across the condenser rises linearly with time according to the equation $V = Cit$. The second tube T_2 is filled with gas and has the property of passing current only when the voltage across it reaches a certain value, which in turn is controlled by the grid potential. When the condenser volt-

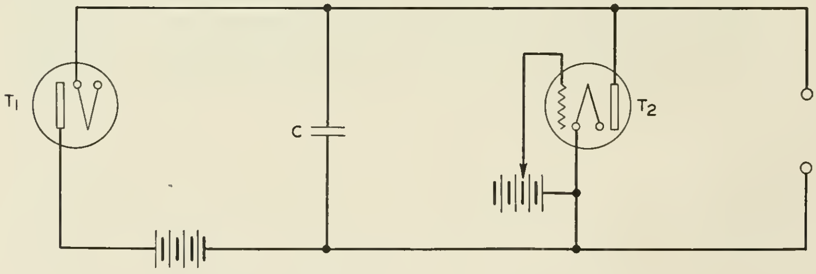


Fig. 22—Thermionic tube "sweep circuit."

---

[21] A. L. Samuel, *Rev. Sci. Inst.*, 2, p. 532, 1931.

age reaches this value the tube operates to discharge the condenser suddenly, and then the uniform charging process begins over again. The condenser voltage applied to one pair of plates of the oscillograph tube makes the spot travel over the screen at a uniform rate in one direction, returning much faster in the other direction. This process may be repeated once a second, or many thousand times a second, depending on the frequency of the wave that is being studied.

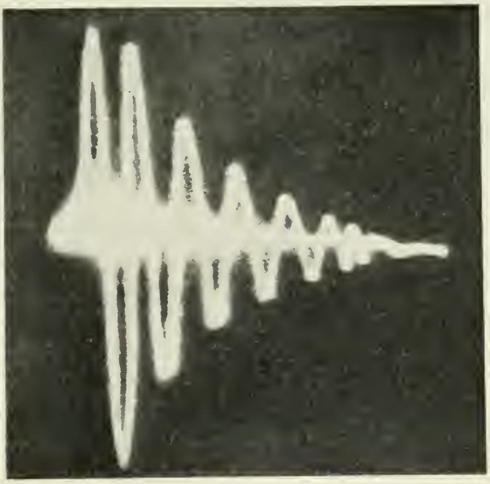


Fig. 23—Discharge of condenser through inductance.

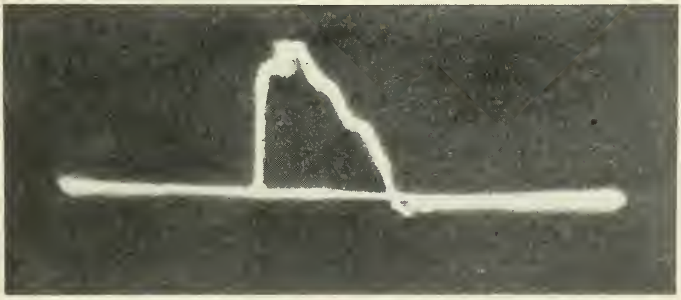


Fig. 24—Discharge of condenser through chattering contact.

In Figs. 23–29 are shown a number of records plotted on a time basis. Figs. 23–25 were made with the Western Electric tube, Figs. 26–27 with the Dufour oscillograph, and Figs. 28–29 on an oscillograph of the type of Rogowski.

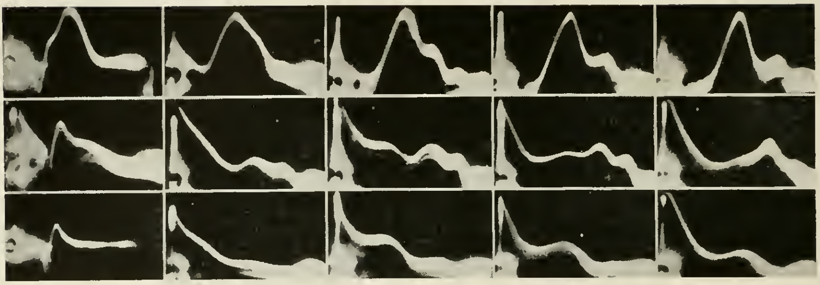


Fig. 25—Action current of a stimulated frog-nerve.[22]

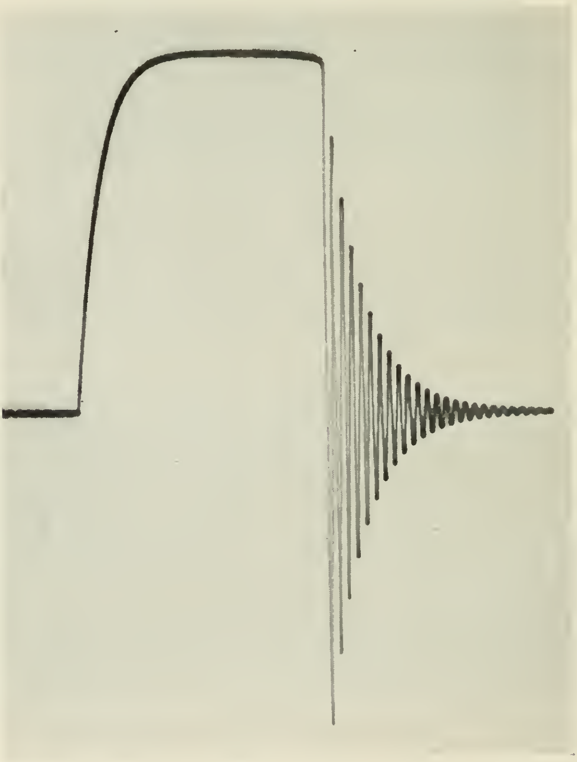


Fig. 26—Make of current through direct contact and break through tuned circuit—Dufour.[14]

---

[22] Gasser and Erlanger, *Am. Jl. Physiol.*, 73, p. 613, 1925.

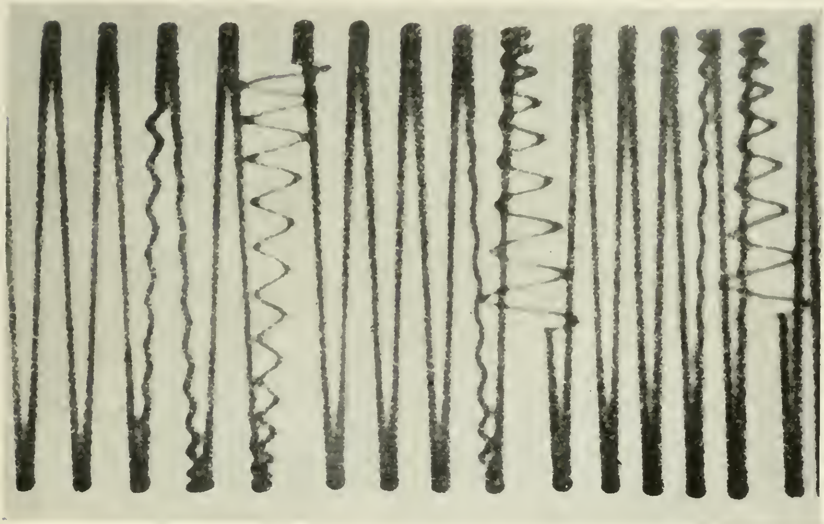


Fig. 27—Wave shape at 8,500,000 cycles—Dufour.[14]

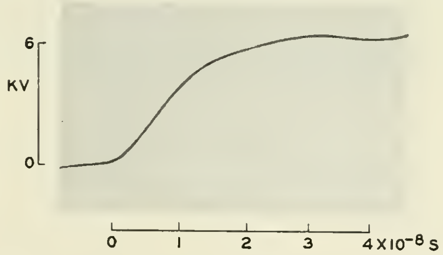


Fig. 28—Front of a voltage wave traveling on a conductor.[23]

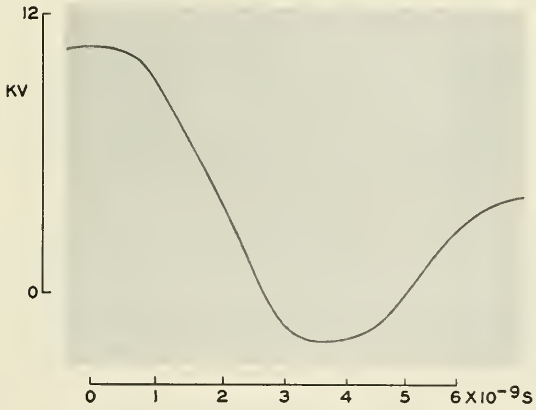


Fig. 29—Initiation of spark in a gas.[24]

[23] Krug, W., *E. T. Z.*, **51**, p. 605, 1930.
[24] Krug, W., *Zeits. f. techn. Phys.*, **11**, p. 153, 1930.

Another way to use the tube is to have it plot the relation between two quantities irrespective of time. As a simple case we may take the current *vs.* voltage curve of a resistance through which an alternating current is flowing. The spot travels back and forth along a straight line, the slope of which measures the reciprocal of the resistance (Fig. 30a). If inductance is added to the resistance the spot does not come down on the same line it went up; an ellipse results, the spread of which tells us the amount of the inductive impedance (Fig. 30b). This method of operation has a wide variety of applications. Instead of the resistance, for instance, we may have a gas discharge device of which we want to know the properties. One of the applications of this method is the production of hysteresis curves of the ferromagnetic materials[25] (Fig. 31). Fig. 32 illustrates the application of the method to the study of distortion in an amplifier.

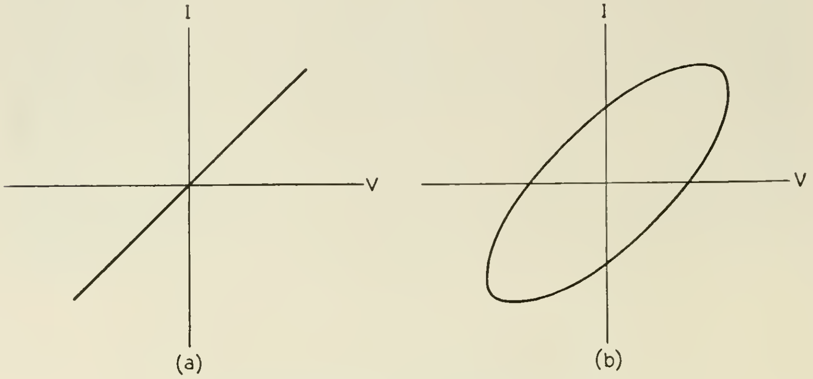


Fig. 30—Diagrams of current voltage curves.

Suppose we apply to each pair of deflector plates a voltage from each of two different oscillators. If the oscillators make exactly the same number of vibrations per second, then the pattern on the tube remains stationary, but if the frequency of the oscillators differ ever so little the pattern goes through gradual changes according to the different phase relations. This is one of the most sensitive means we have for comparing and calibrating accurate oscillators, and we may call it the Lissajous figure method. Fig. 33 shows the appearance of some of these stationary Lissajous patterns.

Another method of comparing the frequency of oscillators has been called the gear-wheel method. In this method the voltage from the low frequency source is split into two equal components 90° apart in

---

[25] Johnson, J. B., *Bell System Technical Jl.*, 8, p. 286, 1929.

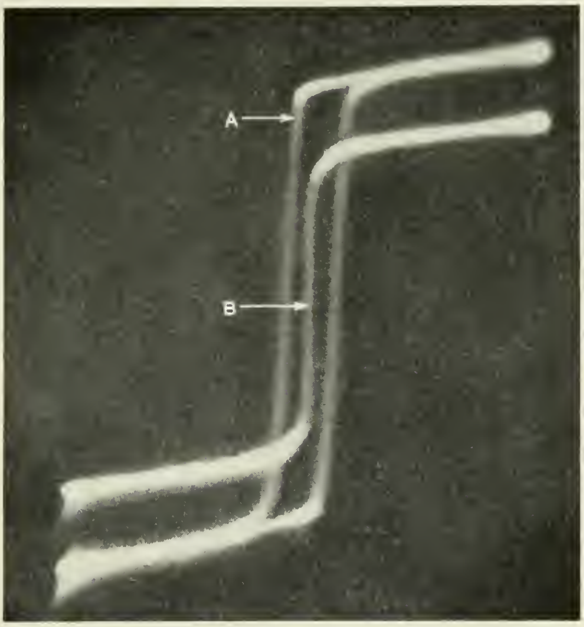


Fig. 31—Hysteresis curves (*a*) of iron, (*b*) of permalloy.

phase. These two voltages are applied to the deflector plates, producing a stationary circle on the screen. The voltage from the oscillator of higher frequency is introduced in the circuit between the cathode and anode of the tube so that the sensitivity of the tube is varied in accordance with the higher frequency. The circle is distorted into a gear-wheel shape as shown in Fig. 34, provided that the

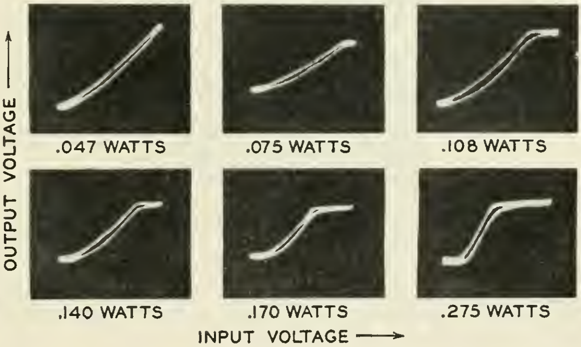
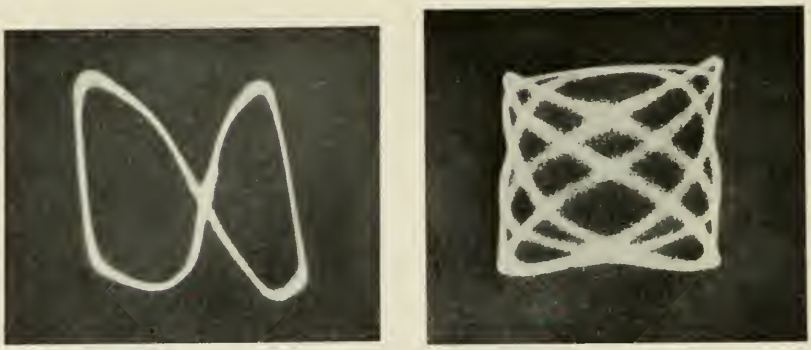


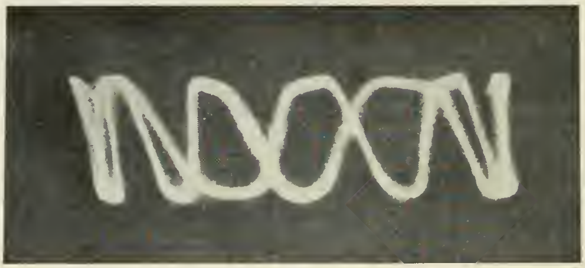
Fig. 32—Distortion in vacuum tube amplifier.[26]

---

[26] Willis and Melhuish, *Bell System Technical Jl.*, **5**, p. 573, 1926.



(*a*) Ratio 2 : 1. (*b*) Ratio 5 : 4.



(*c*) Ratio 8 : 1.

Fig. 33—Frequency comparison.

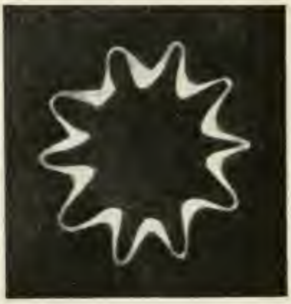


Fig. 34—"Gear wheel" frequency comparison, ratio 10 : 1.

higher frequency is an exact multiple of the lower. If the frequency ratio is not a rational number the gear-wheel rotates, showing the amount of the lack of synchronization.

Two interesting applications to the study of radio telephony are illustrated in Figs. 35 and 36. Fig. 35 shows two characteristics of a transatlantic radio channel, the tube plotting the amplitude of the received signal at a number of modulating frequencies. In Fig. 36 are plotted the magnitude and direction of static crashes as they were shown on the screen of the oscillograph tube.

For demonstration purposes the tube may be used to determine the value of the ratio of charge to mass of the electron, designated by e/m. The classical method is obvious from the derivation of the sensitivity of the tube given above. Especially is this so if we can assume the two fundamental relations $\frac{1}{2}mv^2 = eV$; $mv = eHR$. The method is subject to some error because the extent of the deflecting fields cannot be exactly determined, and because of the space charges in the tube.

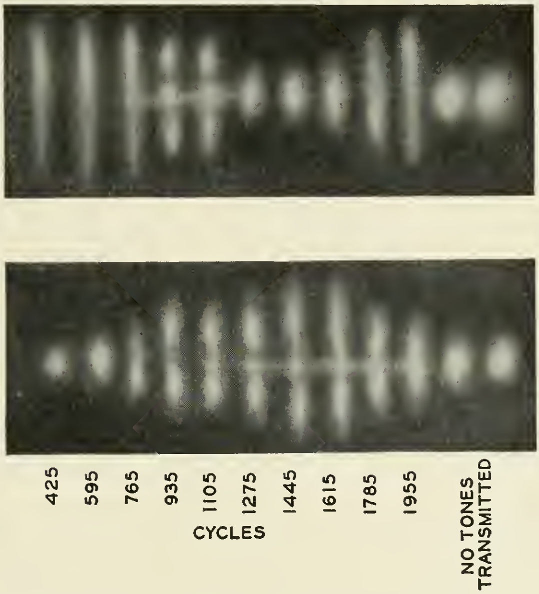


Fig. 35—Transmission characteristic of transatlantic short-wave radio channel.[27]

A more accurate method of arriving at the value of e/m is due to H. Busch.[28] A long solenoid carrying constant current creates a uniform magnetic field in the tube parallel to its axis. The electrons travel in spirals in this field in such a way that when the field has one of a series of values the electrons are focused on the screen. These critical values of the magnetic field are given by the equation

$$H = \frac{2\pi}{L}\sqrt{\frac{2m}{e}}\,V;$$

---

[27] Potter, R. K., *Inst. Radio Eng.*, **18**, p. 581, 1930.
[28] Busch, H., *Phys. Zeits.*, **23**, p. 438, 1922.

where $n = 1, 2, 3.$—etc., L is the length of the beam and V the driving potential.

In this method there is less disturbance from gas focusing and space charge if the filament current is made as low as will give a still visible spot with the magnetic focusing.

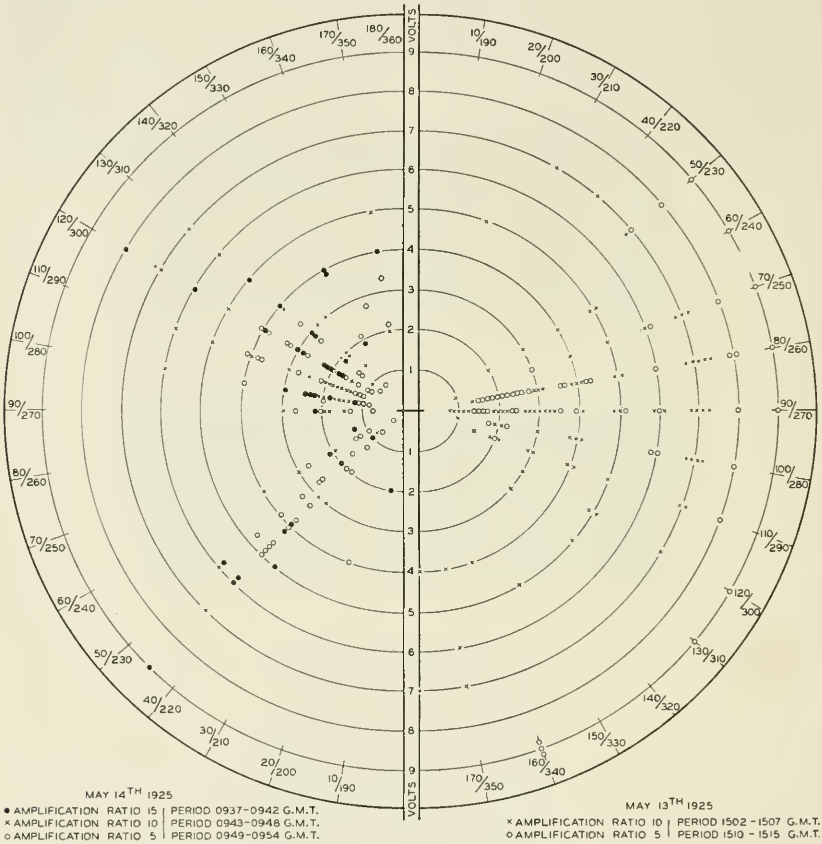


Fig. 36—Azimuthal distribution of atmospherics.[29]

From the beginning the cathode ray oscillograph was recognized as a tool of great promise. Its use was limited, however, by the difficulties of maintaining a constant degree of vacuum and of providing a suitable source of high voltage, as well as by the general bulk and clumsiness of the apparatus. We have seen that in the nearly thirty-five years since the first invention of the tube, many improvements have been made both in the structure of the tube and in the methods

[29] Watson Watt and Herd, *Jl. I. E. E.*, **64**, p. 611, 1926.

of operation so that the handicaps under which the early apparatus was used have been very largely eliminated. The introduction of the low voltage, high sensitivity tube of moderate cost expanded the use of cathode ray oscillographs rapidly, until now they are used in almost every laboratory where high frequency measurements are made. In the Bell System alone more than a hundred of these instruments are in constant use for research and for the control and calibration of manufactured products. Originally an intractable and little used device, the cathode ray oscillograph has become an almost universal scientific and industrial tool.

The Operation of Vacuum Tubes as Class B and Class C Amplifiers *

By C. E. FAY

A simple theoretical development of the action of a vacuum tube and its associated circuit when used as a Class B or Class C amplifier is given. An expression for the power output is obtained and the conditions for maximum outputs are indicated. The way in which the tuned plate circuit filters out the harmonics in the pulsating plate current wave is illustrated by an hypothetical example. A set of dynamic output current characteristics is developed graphically from a set of static characteristics. The Class B dynamic curves are found to give a better approximation to a straight line than the Class C curves because of a reversed curvature which appears at the lower ends. It is pointed out that the screen grid tube should function similarly to a high μ three-element tube in this type of operation. Experimental dynamic characteristics of a three-element tube, Western Electric 251-A, and of a screen grid tube, Western Electric 278-A, of identical dimensions are shown which verify the theoretical results. The screen grid tube gives about the same output and efficiency as the three-element tube, but its dynamic characteristic tends to bend more rapidly at the upper end.

INTRODUCTION

THE majority of modern radio telephone installations in this country are designed for modulation at a low power level and amplification of the modulated carrier, so as to obtain up to 100 per cent modulation in the output stage. As far as the writer is aware there seems to be no very comprehensive material dealing with this phase of vacuum tube operation available from past publications. A treatment of the operation of such amplifiers seems particularly desirable from the standpoint of the design of vacuum tubes for such service.

Some of the fundamental considerations regarding Class B or C operation were given by Morecroft and Friis,[1] and later, a more complete analysis of power oscillators by Prince.[2] Both of these, however, were primarily concerned with the attainment of steady output at high efficiency. Other papers [3, 4, 5] of more recent date have touched somewhat upon the subject.

---

* To appear in *Proc. I. R. E.*, March, 1932.

[1] J. H. Morecroft and H. T. Friis, "The Vacuum Tube as a Generator of Alternating Current Power," *Transactions A. I. E. E.*, Vol. 38, No. 2, Oct., 1919.

[2] D. C. Prince, "Vacuum Tubes as Power Oscillators," *Proc. I. R. E.*, Vol. 11, Nos. 3, 4, 5, June, Aug., Oct., 1923.

[3] A. A. Oswald and J. C. Schelleng, "Power Amplifiers in Transatlantic Radio Telephony," *Proc. I. R. E.*, Vol. 13, No. 3, June, 1925.

[4] E. E. Spitzer, "Grid Losses in Power Amplifiers," *Proc. I. R. E.*, Vol. 17, No. 6, June, 1929.

[5] Y. Kusunose, "Calculation of Characteristics and the Design of Triodes," *Proc. I. R. E.*, Vol. 17, No. 10, October, 1929.

This paper will deal particularly with the type of amplifier used for the amplification of the modulated carrier. It will be assumed that a linear relation between input voltage and output current is the desired characteristic of such amplifiers and that the more nearly this relation is attained, the less will be the distortion produced.

An approximate graphical method for the calculation of the dynamic output characteristic from the static characteristics of a tube is outlined which is capable of considerable accuracy. The exciting voltage is taken to be a sinusoidal voltage of varying amplitude, and only the fundamental component of the output current is considered. The very important question of the distortion introduced by the non-linearity of the characteristic and by the resonance effect of the output circuit, as well as the question of the suppression of harmonics in the antenna circuit, is considered to be beyond the scope of this paper.

Theoretical Development

Class B [6] amplifiers have been defined as those which operate with a negative grid bias such that plate current is practically zero with no excitation grid voltage, and in which the power output is proportional to the square of the excitation voltage.

Class C amplifiers have been defined as those which operate with a negative grid bias more than sufficient to reduce the plate current to zero with no excitation grid voltage, and in which the output varies as the square of the plate voltage between limits.

There is actually very little distinction between the two types as the fundamental principles of operation are the same in that the plate current flows in pulses and becomes zero during part of the cycle, the Class C type being merely the case where the duration of the pulses is shorter. For the purposes of this paper, a Class B amplifier shall be regarded as one in which the grid bias is either just sufficient or is not sufficient to reduce the plate current to zero with no excitation grid voltage, and a Class C amplifier as one in which the grid bias is more than sufficient to reduce the plate current to zero with no excitation grid voltage.

Let us consider the plate circuit of a Class B or C amplifier to be represented schematically in Fig. 1. We will omit the grid circuit and assume that the excitation of the grid merely varies the internal tube resistance, R_p, which it does in effect. If at the start, the grid is so biased that no plate current flows, the condenser C_0 will charge up to a potential E_b as indicated in Fig. 1. Let it be assumed that C_0 is of sufficiently large capacity that it presents negligible impedance at the frequency of operation, or that the time constant of C_0 and R_0 is large compared to the time of one cycle of the operating frequency. In this event, then, the voltage across C_0 will remain constant at E_b during a complete cycle. Let it also be assumed that the choke coil L_b is of sufficient inductance that the current I_b is maintained constant throughout the cycle.

---

[6] 1931 Standardization Report, Year Book of the I. R. E., 1931, p. 71.

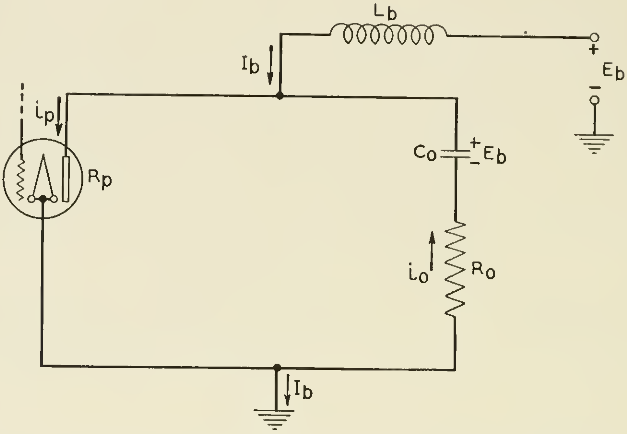


Fig. 1—Schematic of amplifier plate circuit with resistance load.

Then by applying Kirchoff's laws to the circuit, Fig. 1, remembering the assumptions regarding C_0 and L_b, we find the following relations must hold at any instant:

$$E_b = i_p R_p + i_0 R_0 \tag{1}$$

or

$$e_p = E_b - i_0 R_0 \tag{2}$$

and

$$i_0 = i_p - I_b, \tag{3}$$

also

$$I_b = \text{Average of } i_p \text{ over 1 cycle}, \tag{4}$$

since the average current through C_0 must be zero.*

Then let it be assumed that the grid of the tube is excited and biased in such a manner that the plate current, i_p, will vary sinusoidally as illustrated in Fig. 2. When the resistance R_p is at its minimum value, i_p will be a maximum and will be the sum of I_b and i_0, [(1) and (3)]. Also e_p will be a minimum, (2); see point 1, Fig. 2. As R_p is increased by the grid potential going in the negative direction, i_p will be reduced

* See Table of Symbols at end of this paper.

as also will i_0. At the instant when

$$i_p = I_b, \qquad i_0 = 0,$$

as illustrated by point 2, Fig. 2. Then as R_p is further increased, i_p becomes less and i_0 starts to increase in the negative direction until i_p is zero at cut-off;

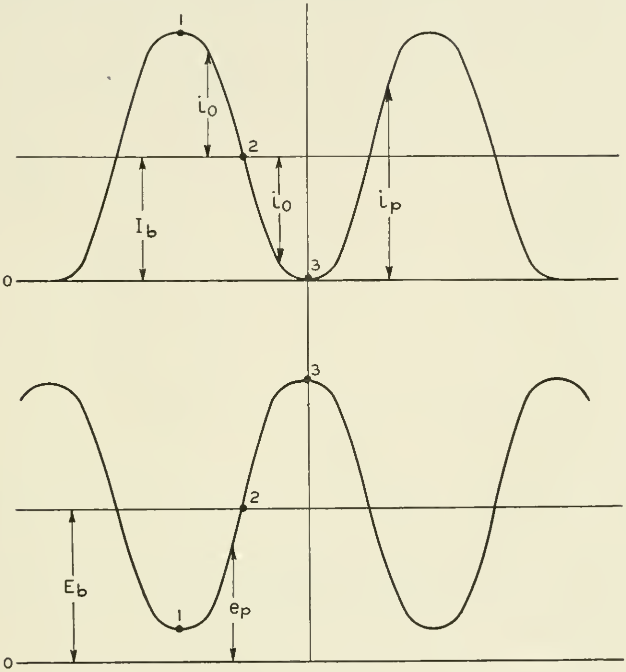


Fig. 2—Plate current and plate voltage relations for sinusoidal plate current in circuit of Fig. 1.

$$R_p = \infty, \qquad i_0 = -I_b \text{ from (3).}$$

Then from (2) and the above,

$$e_p = E_b + I_b R_0, \tag{5}$$

which is the peak value of e_p. Thus the power in R_0 is $i_0^2R_0$ averaged over a cycle and the power dissipated in the tube is $i_p^2R_p$ averaged over a cycle.

In actual operation, the plate current is not sinusoidal but goes from zero to peak value and back to zero in about half the cycle and remains cut off during the rest of the cycle, as illustrated by Fig. 4A. Also, instead of being a pure resistance, the output circuit consists of a tuned tank circuit, Fig. 3, into which resistance is introduced either directly or by coupling in some way. Thus the impedance of the output circuit will be a resistance to the fundamental frequency only. However, the same general principles will apply in this case as in the case of the sinusoidal plate current and the pure resistance circuit. The equations will be of the same form except that Z_0 must be substituted for R_0 as follows:

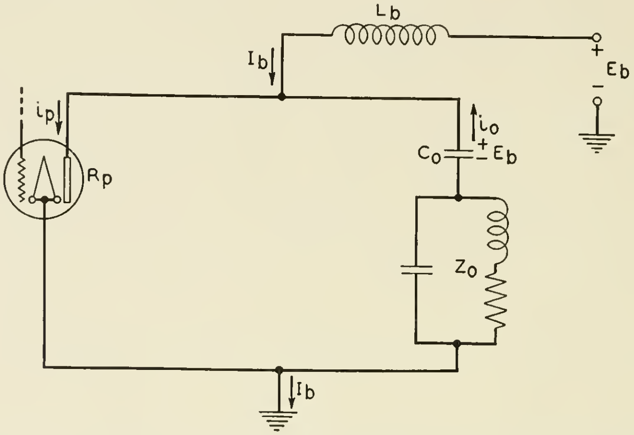


Fig. 3—Schematic of amplifier plate circuit with tuned output impedance.

$$E_b = i_p R_p + i_0 Z_0, \tag{1A}$$

$$e_p = E_b - i_0 Z_0, \tag{2A}$$

$$i_0 = i_p - I_b, \tag{3}$$

$$I_b = \text{Average of } i_p \text{ over 1 cycle.} \tag{4}$$

In this case it must be remembered that the wave of i_p, and hence i_0, instead of being a simple sine wave, is a more complex wave consisting of a fundamental frequency and numerous harmonics. Also $i_0 Z_0$ is the sum of all such components multiplied by the respective impedances presented to them taken in their proper phases.

Considering the wave form of i_0 shown in Fig. 4B, which is readily obtainable in practice, it should be evident from inspection that it can be considered a cosine wave containing odd and even cosine terms and may be expressed by

$$i_0 = I_1 \cos \omega t + I_2 \cos 2\omega t + I_3 \cos 3\omega t + I_4 \cos 4\omega t + \dots + I_n \cos n\omega t.$$

By means of an harmonic analysis of this wave for the numerical values indicated in Fig. 4B, the coefficients were found to be approximately as follows:

$$I_1 = 0.96, \quad I_2 = 0.543, \quad I_3 = 0.140, \quad I_4 = -0.07, \quad I_5 = -0.105, \quad I_6 = -0.043.$$

The voltage produced across the output circuit by the wave, i_0, then will be the sum of the voltages produced by each one of the com-

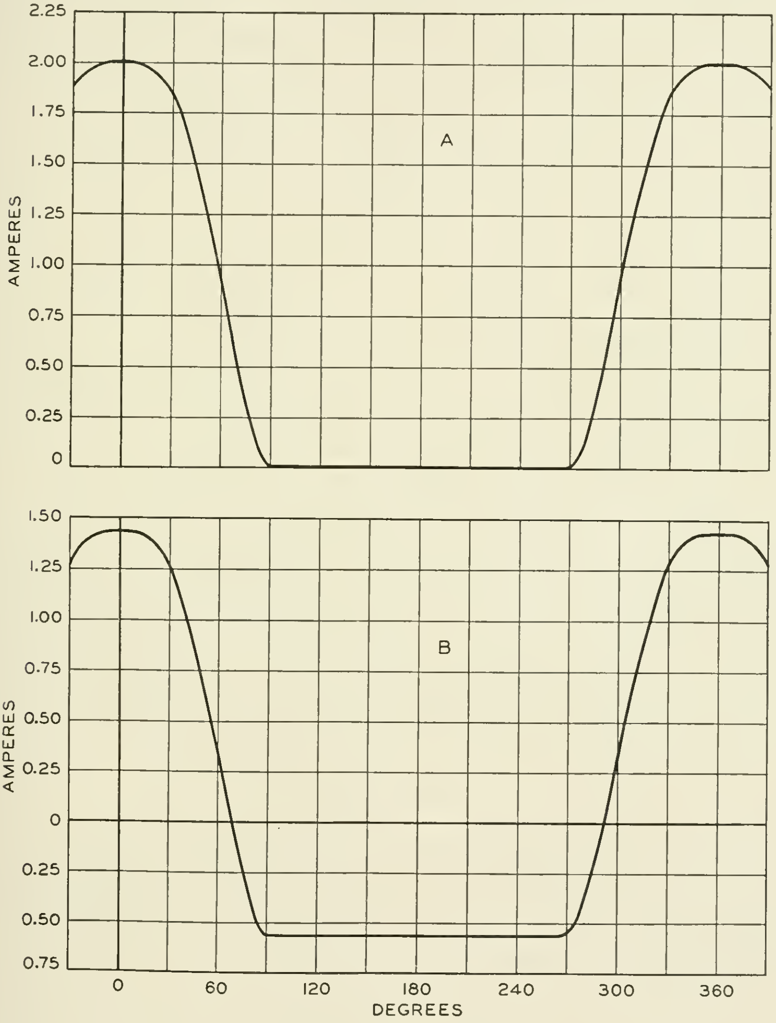


Fig. 4—*A*. Pulsating plate current wave, i_p. *B*. Alternating component of plate current, i_0.

ponents of i_0. Let it be assumed that the reactances of the output circuit are 300 ohms each at fundamental frequency and that sufficient resistance has been inserted in the inductive branch to make the impedance to fundamental frequency 2000 ohms resistance. This will require about 45 ohms in series with the inductance. At second harmonic frequency, then, the impedance of the inductive branch in complex notation is $45 + j600$ and that of the capacity branch $0 - j150$. Without introducing much error in the result we may as well write $+ j600$ and $- j150$, which in parallel give $- j200$. At third harmonic frequency we have $+ j900$ and $- j100$ which in parallel give $- j112.5$. At fourth harmonic frequency we have $+ j1200$ and $- j75$ which in parallel give $- j80$.

Thus the voltage produced by the fundamental will be

$$e_1 = 0.96 \times 2000 \cos \omega t = 1920 \cos \omega t,$$

and that produced by the second harmonic will be

$$e_2 = 0.54 \times 200 \cos (2\omega t - 90^\circ) = 108 \cos (2\omega t - 90^\circ),$$

and that produced by the third harmonic will be

$$e_3 = 0.14 \times 112.5 \cos (3\omega t - 90^\circ) = 15.75 \cos (3\omega t - 90^\circ),$$

and that produced by the fourth harmonic will be

$$e_4 = - 0.07 \times 80 \cos (4\omega t - 90^\circ) = 5.6 \cos (4\omega t + 90^\circ),$$

etc. Fig. 5B shows to scale the fundamental and second harmonic voltages produced, and the dotted curve gives the sum. The higher harmonic voltages are too small to show on the plot. The resultant is seen to be very little different from a sine wave. Therefore, even though the wave i_0 departs radically from a sine wave, the voltage produced across the output circuit is very nearly sinusoidal. Then the output power at fundamental frequency is

$$W_0 = \frac{I_1{}^2 R_0}{2}, \tag{6}$$

where I_1 is the peak value of the fundamental component of i_0, and R_0 is the effective resistance of the tank circuit at fundamental frequency.

The output power may also be expressed:

$$W_0 = \frac{(E_b - e_{pm}) K I_p}{2} \tag{7}$$

if $I_1 = KI_p$.

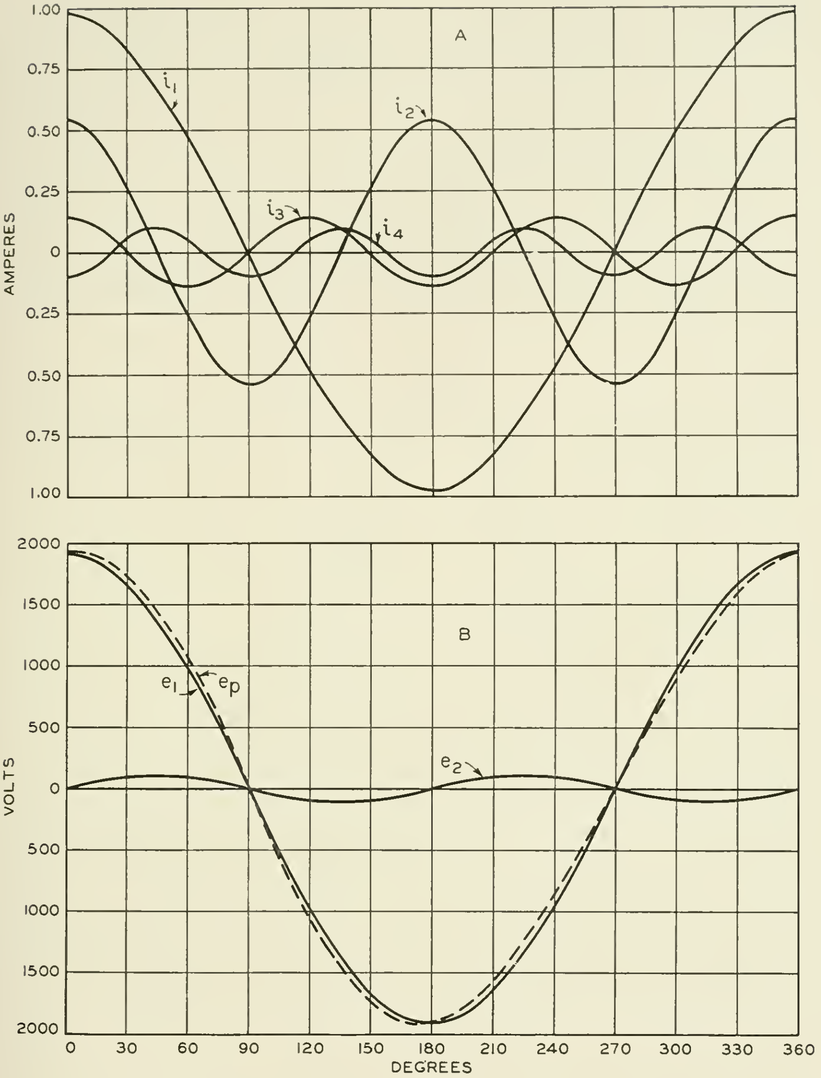


Fig. 5—*A*. First four components of i_0, Fig. 4*B*. *B*. Alternating components of plate voltage produced by i_0 in output circuit.

For any constant value of grid voltage, the plate current will be some function of the plate voltage, so that for any peak value of e_g we may write

$$e_{pm} = f(I_p). \tag{8}$$

Substituting (8) in (7) we have

$$W_0 = \frac{[E_b - f(I_p)]KI_p}{2}. \tag{9}$$

In order for W_0 to be a maximum,

$$\frac{dW_0}{dI_p} = 0.$$

Performing this operation on (9) and solving for I_p assuming K is constant, we get

$$I_p = \frac{E_b - f(I_p)}{f'(I_p)} = \frac{E_b - e_{pm}}{r_p} \tag{10}$$

since $f'(I_p)$ is $df(I_p)/dI_p$, which is obviously r_p, the differential plate resistance when I_p is flowing.

We may also write

$$R_0 = \frac{E_b - e_{pm}}{KI_p}. \tag{11}$$

Substituting (10) in (11) we obtain

$$R_0 = \frac{r_p}{K} \tag{12}$$

which gives the relation of R_0 to r_p for maximum power output.

If we have the $i_p - e_p$ curves for any tube we may approximate closely the point of maximum output at any peak grid voltage. This can be accomplished by a process of cut and try in finding where the quantity $(E_b - e_{pm})I_p$ becomes a maximum, since K remains fairly constant, depending mostly on the bias voltage and grid excitation voltage, as will be shown later. Also, for any given output impedance and peak grid voltage, we may find the output from the $i_p - e_p$ curves by cut and try by finding where

$$\frac{E_b - e_{pm}}{I_p} = KR_0 \qquad \text{[from (11)]}$$

is satisfied. It will be noticed from Figs. 4A and 5A that $I_1/I_p = 0.48$. In general, K will be near this value for well loaded conditions. The actual value depends upon the exciting voltage and grid bias voltage, and at low values of excitation it may differ considerably from the value indicated above which may render the cut and try methods outlined subject to considerable error.

In addition to finding the maximum output for a given set of conditions it is desirable to know what the shape of the curve of output current versus exciting voltage will be. The following method determines points on the output curve by the use of the static characteristic of the tube.

In Fig. 6 is shown a family of plate current curves obtained by applying a sine wave grid voltage to a three-halves power characteristic with a plate voltage consisting of a steady voltage plus a sine wave 180° out of phase with the grid voltage. The different curves show the relative shapes obtained by variation of the grid bias so that the portion of the cycle during which plate current flows is varied. It can be shown that for any tube, no matter what the actual values of voltages and currents are, as long as the portion of the characteristic under consideration obeys the three-halves power law, the plate current waves will correspond to those of Fig. 6 in shape for the same respective periods of plate current flow. By means of harmonic analyses, the value of *K* corresponding to each of these shapes was calculated and Fig. 7 shows the variation of *K* with the number of degrees during which plate current flows. Fig. 8 shows a comparison of three possible shapes for 180° flow with the corresponding values of *K*. *A* is the unsaturated curve of Fig. 6. *B* shows a curve for which the characteristic departs from the three-halves power law at its upper end, presumably due to the effects of grid current, etc. This curve if unsaturated would have a peak value about 17 per cent higher. *C* shows

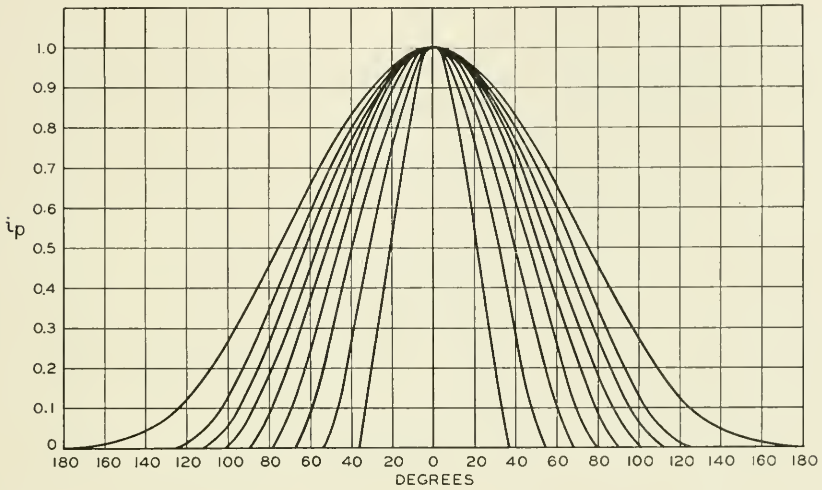


Fig. 6—Curves of i_p obtained from a three halves power characteristic with sinusoidal exciting voltage for varying periods of plate current flow.

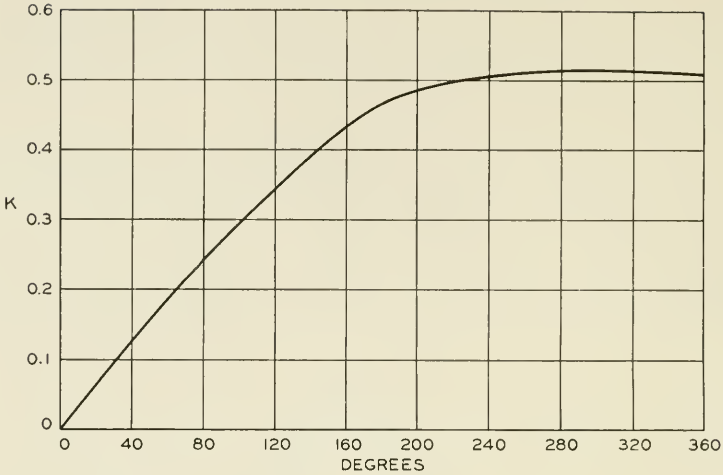


Fig. 7—Variation of *K* with period of flow for curves of Fig. 6.

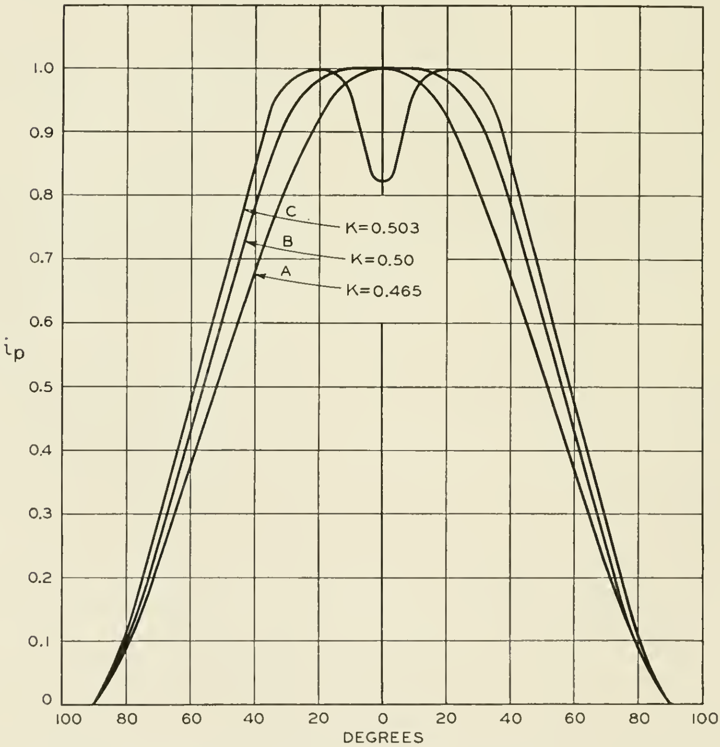


Fig. 8—Three typical shapes of plate current for 180° flow. *A*, unsaturated. *B*, slightly saturated. *C*, extremely saturated.

a curve for which the saturation is so pronounced that the grid current drawn has caused a depression.

In actual tubes the static characteristics do not follow the three-halves power law exactly, but for the purposes of this paper it is sufficiently accurate to assume that they do in the portion where grid current is not appreciable.

With the information available from Figs. 6 and 7, and the static characteristic of a tube, we should be able to plot the dynamic characteristic for any value of output impedance. For an example, let Fig. 9 represent the static characteristics of a tube (Western Electric No. 251-A). Then for some value of output impedance, say 2000 ohms, let us plot curves of constant K on the static characteristic assuming $E_b = 3000$ volts. Points on these curves are found from

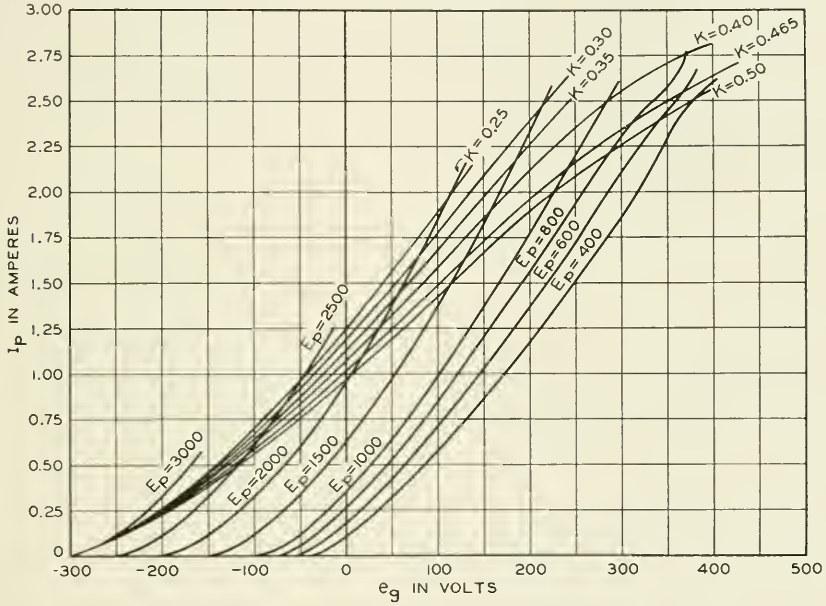


Fig. 9—Static characteristic with lines of constant K for 2000 ohms impedance. (Western Electric No. 251-A Tube.)

$$I_p = \frac{E_b - e_{pm}}{KR_0},$$

which is another form of (11). For example, to find where the $K = 0.5$ curve crosses the $E_p = 2000$ curve, we have $e_{pm} = E_p = 2000$, $E_b - e_{pm} = 1000$, thus

$$I_p = \frac{1000}{0.5 \times 2000} = 1.0 \qquad \text{(Fig. 9).}$$

The cut-off point, or the potential which the grid must have to make the plate current just zero is given by

$$e_g \text{ cut-off} = -\frac{E_b}{\mu}. \tag{13}$$

If we bias the grid with this negative voltage, it may be said to be biased at cut-off, and if the grid voltage becomes more positive than this value, plate current will flow. Thus if we apply a sinusoidal exciting voltage, the plate current will always flow during 180°, or half the cycle, and for this condition $K = 0.465$ from Fig. 7, so the dynamic output current curve will be obtained by multiplying the ordinates of the $K = 0.465$ curve of Fig. 9, by 0.465. However, in the upper portion where the static characteristics depart appreciably from the three-halves power law, K will increase gradually (Fig. 8), so that it is more accurate to use slightly increasing values of K in this portion. (See table of calculations for $E_c = 300$, Fig. 10.)

$E_b = 3000$ Volts $Z_0 = 2000$ ohms

E_c Volts	Peak Excitation Volts	Peak e_g Volts	K	I_p Amp.	I_1 Amp.	$\frac{I_1}{\sqrt{2}}$ Amp.	Watts Output $\left(\frac{I_1}{\sqrt{2}}\right)^2 Z_0$	Output Current in 150 ohm Ant. Amp.
−300	100	−200	0.465	0.230	0.107	0.0756	11.5	0.276
	200	−100	0.465	0.575	0.267	0.1885	71	0.688
	300	0	0.465	1.01	0.47	0.332	222	1.218
	400	+100	0.465	1.49	0.692	0.490	480	1.790
	500	200	0.47	1.95	0.917	0.647	840	2.37
	600	300	0.48	2.30	1.07	0.756	1150	2.96
	650	350	0.49	2.45	1.20	0.848	1440	3.10
−350	100	−250	0.345	0.112	0.039	0.027	1.49	0.099
	200	−150	0.415	0.410	0.170	0.122	28.9	0.438
	300	− 50	0.435	0.810	0.352	0.249	124	0.907
	400	+ 50	0.445	1.280	0.570	0.403	324	1.470
	500	150	0.45	1.77	0.797	0.563	635	2.055
	600	250	0.452	2.20	0.995	0.702	988	2.56
	650	300	0.453	2.40	1.087	0.768	1180	2.80
−250	50	−200	0.51	0.220	0.112	0.079	12.6	0.290
	100	−150	0.505	0.375	0.190	0.134	36	0.490
	200	− 50	0.492	0.760	0.374	0.265	140	0.966
	300	+ 50	0.483	1.220	0.590	0.417	348	1.525
	400	150	0.480	1.710	0.820	0.580	674	2.120
	500	250	0.477	2.137	1.020	0.721	1040	2.635
	600	350	0.475	2.475	1.173	0.830	1380	3.030

Fig. 10—Calculation of dynamic output characteristics—No. 251-A tube.

If the grid bias is more negative than the cut-off point (i.e., if the tube is biased "below cut-off"), the plate current will not begin to flow until the grid potential has reached the cut-off point (13), and the portion of the cycle in which plate current flows will be determined by the number of degrees of the cycle in which the exciting wave is above the cut-off point. Thus in this case, K will depend on the amplitude of the exciting wave compared to the difference between the bias and the cut-off point. When the amplitude becomes large compared to this difference, the period of plate current flow approaches 180° but when the peak value of the exciting wave just reaches the cut-off point, the period of flow is zero. In this case K can vary from almost 0.465 to zero depending on the amplitude of the exciting wave, or if the exciting voltage is sufficient to produce a saturation effect in the plate current wave, K might exceed 0.465. This is the case of the Class C amplifier.

If the grid bias is more positive than the cut-off point (i.e., if the tube is biased "above cut-off"), plate current flows during 360° of the cycle until the amplitude of the exciting voltage is sufficient to reach the cut-off point. The period of flow continues to decrease until it approaches 180° as the amplitude of the exciting voltage becomes large compared to the difference between the bias and cut-off potentials. Here K is always greater than 0.465; see Fig. 7. This is the case of the Class B amplifier.

Fig. 10 outlines the calculation of dynamic characteristics of both types from the static characteristics of Fig. 9. In these calculations, the cut-off point is assumed to be constant at -300 volts and the value of K is obtained from Fig. 7 after determining the portion of the cycle in which the exciting voltage is above the cut-off point. For $E_c = -300$ of course it is always 180°. For $E_c = -350$, and a peak exciting voltage of 100 volts for example, the peak e_g will be -250 which will give 120° for the period of flow and thus $K = 0.345$ from Fig. 7. Then on the static characteristic, Fig. 9, taking $e_g = -250$, and by interpolating between the $K = 0.30$ and $K = 0.35$ curves to $K = 0.345$ we find $I_p = 0.112$. Multiplying this by K we get $I_1 = 0.0386$. Changing to r.m.s. value, squaring and multiplying by 2000 ohms, we obtain 1.49 watts for the output power. It is then determined that this would represent a current of 0.0995 ampere in the 150-ohm dummy antenna which was used in the experimental work, and which was so coupled to the output tank circuit that the impedance into which the tube was working was of the value indicated.

In the calculations of Fig. 10, the cut-off point was assumed constant at $-E_b/\mu$. However, it actually varies as $-e_p/\mu$ and will vary sinusoidally if e_p is sinusoidal. No error is introduced by this fact in the case of bias at cut-off ($E_c = -300$). However, in the other two cases an error is introduced. In the case for bias below cut-off ($E_c = -350$) the amount of error is indicated by the two points illustrated in Fig. 11. *A* is the case for a peak exciting voltage of 300 volts. From the static characteristics of Fig. 9 we find that e_{pm} will be about 2300 volts which will make -230 volts, the peak of the real cut-off curve ($-e_p/\mu$) and since it is a sine wave we can plot

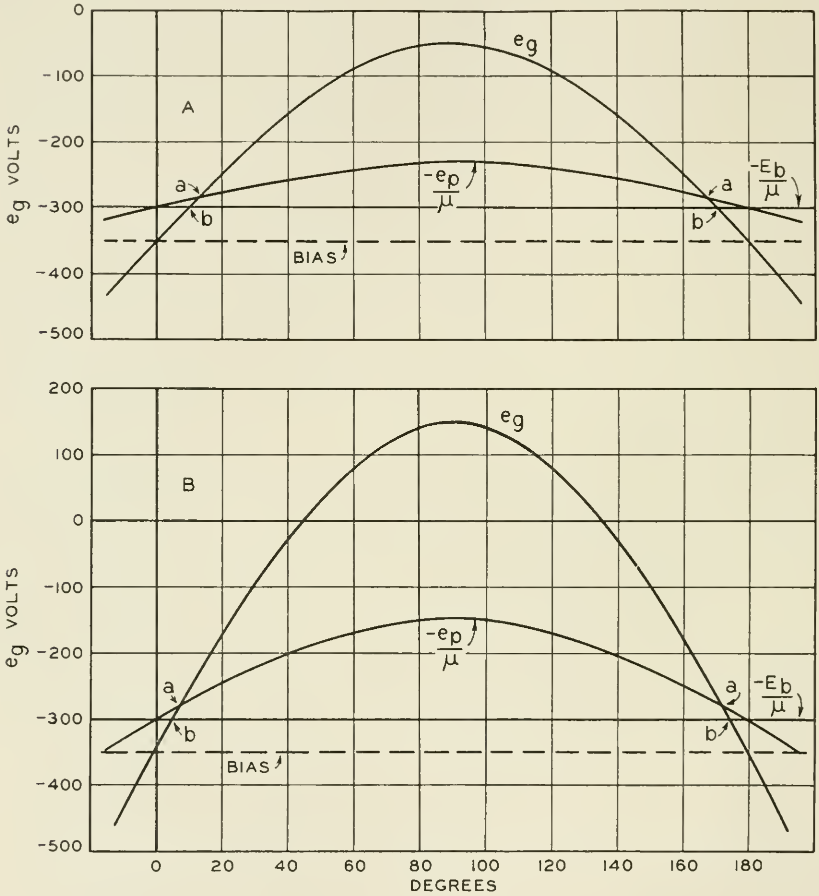


Fig. 11—Illustrating method of obtaining actual period of flow of plate current. *a*, acutal cut-off point, *b*, cut-off point assumed in calculations.

it as shown. The error in the period of flow then will be represented by the difference between where the e_g curve cuts the -300-volt line and where it cuts the $-e_p/\mu$ curve. The actual period of flow is about 156° instead of 160° as found before, which would make $K = 0.425$ instead of 0.435. *B* is the case for a peak exciting voltage of

500 volts and by the same procedure we find $K = 0.441$ instead of $K = 0.45$. These errors are quite small and do not alter the result appreciably in these cases.

The effect will be less pronounced, the higher the μ of the tube, and more pronounced the further the bias voltage is moved from the cut-off point. For extreme cases in which more accuracy is desired, K will be determined quite accurately by the second approximation method of Fig. 11. For the case of bias above cut-off where K is always greater than 0.465 the error in determining K will be considerably smaller as K does not vary greatly with the period of flow in this region; see Fig. 7. In the upper portion of the characteristic where departure from the three halves power law becomes appreciable, the increase in K due to this saturation effect which must be estimated,

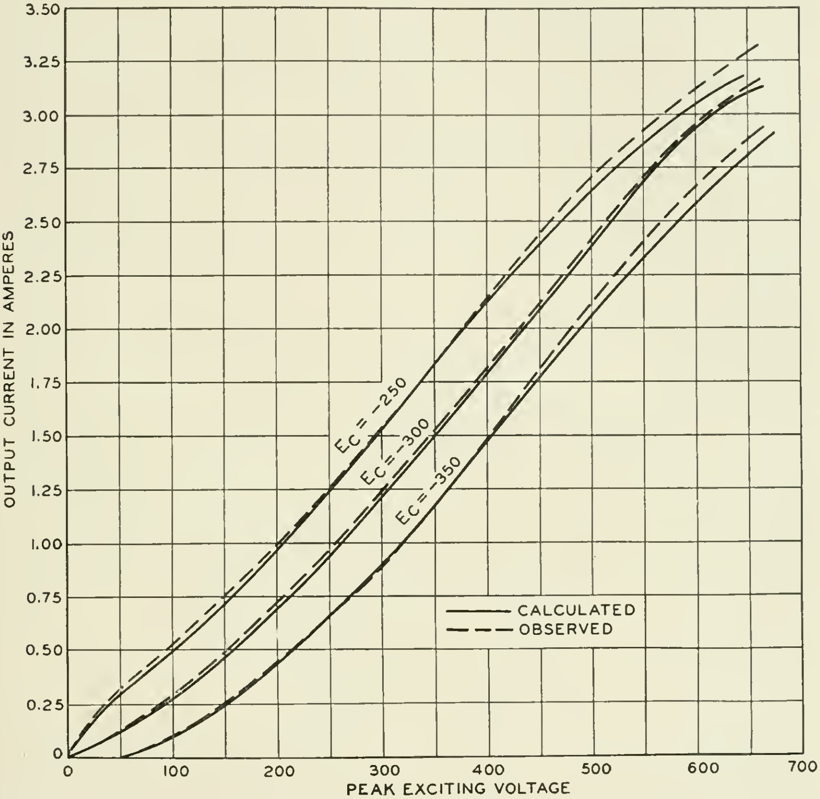


Fig. 12—Calculated and experimental dynamic characteristics of Western Electric No. 251-A Tube $E_b = 3000$ volts, $Z_0 = 2000$ ohms, $\mu = 10$.

Fig. 8, may be greater than the error produced by neglecting the varying cut-off. In this event it would hardly be worthwhile to use the method of Fig. 11, except in extreme cases where the bias differs greatly from the cut-off potential.

The dynamic output current characteristics calculated in Fig. 10 are shown plotted to scale in Fig. 12. It should be noted that for the case of bias above cut-off ($E_c = -250$) the curvature of the dynamic is reversed at the lower end. This will tend to allow a better approximation to a straight line for the overall curve than can be obtained for curves of bias below cut-off.

In general the highest output impedance admissible will give the straightest dynamic curve for a given tube. Comparing two tubes of equal power rating, the one with the highest mutual conductance will give the straightest dynamic characteristic with a given output impedance.

It will be noticed that the dynamic curves tend to show a saturation effect at their upper ends. This is predicted, however, from the static characteristics. If the grid excitation voltage is sufficient to allow the grid potential to approach the plate potential very closely, the electron current taken by the grid will increase rapidly. Since this current taken by the grid would otherwise have been taken by the plate, the result is a reduction in the value of i_p which causes the plate current characteristic to depart from the three-halves power law. A slight saturation effect at the upper end of the dynamic curve may help the curve to approximate a straight line more closely. The output impedance may be chosen so as to realize this advantage provided the grid becomes positive during a sufficient portion of the cycle.

THE SCREEN GRID TUBE

The foregoing theory, with a few alterations, will apply equally well to the screen grid tube. In any screen grid tube where the screening is sufficient to reduce the grid-plate capacity to the point where operation at high frequencies without neutralization is feasible, the screen voltage will determine the plate current at a given grid voltage almost entirely, the plate voltage having very little effect. In this case the cut-off point will be given approximately by

$$e_g \text{ cut-off} = -\frac{E_s}{\mu}, \tag{14}$$

where μ here is the μ of a three-element tube with the plate in place of the screen. A more exact formula would be

$$e_g \text{ cut-off} = -\frac{E_b + \rho E_s}{\mu(1 + \rho)}, \tag{15}$$

where μ is as defined above, and ρ is the amplification factor of a three-element tube considering the screen as the grid. However, if ρ is large compared to μ and μ is greater than 1, the cut-off point is given quite closely by (14).

In operation the screen is by-passed to ground by a large capacity so that the screen potential will remain practically constant at E_s to insure the effectiveness of the screening action. As the minimum plate potential approaches the screen potential, the screen will begin to draw appreciable current and thus cause a saturation effect in the characteristic similar to that of the three-element tube when the minimum plate potential approaches the maximum grid potential. Also, as the maximum grid potential approaches the screen potential, the grid will begin to draw current which will subtract from the plate and screen currents, thus tending also to cause a saturation effect in the characteristic. It would seem desirable, therefore, that the screen voltage be chosen somewhere between the maximum grid potential and the minimum plate potential, the actual point depending on whether grid current or screen current is the least desirable. In operation the screen grid tube should give the same type of characteristics as indicated in Fig. 12 and have some advantage from the fact that as a high μ tube it requires less driving voltage. However, the dynamic characteristic may tend to saturate sooner than it would for the equivalent three-element tube because of the tendency of both the screen and the grid to absorb current under the conditions mentioned above.

Experimental Results

In Fig. 12 along with the curves calculated from the theoretical development are shown the actual experimental curves obtained at 3000 kc for the same conditions.

The following dynamic output current and efficiency curves of the three-element tube, Western Electric 251-A, and the screen grid tube, Western Electric 278-A, measured at 3000 kc are submitted to show the effects of various factors on the dynamic characteristic. They allow a direct comparison of the three-element tube and the screen grid tube.

The curves of Fig. 13 show the dynamic output currents and efficiencies obtained from the three element tube for the conditions indicated. The upper portions of these curves show some saturation effect and they are practically identical except that they are displaced along the abscissa by approximately the differences in grid bias. Fig. 14 shows the effect of varying the output impedance for the case of bias at cut-off. The relative effect for other biases is practically the same as for the case shown.

Fig. 15 shows the dynamic output currents and efficiencies for a screen grid tube which is identical in construction with the three-element tube except for the addition of the screen between the grid and plate. The μ of this tube considering the screen to be the plate is about 4, so that the cut-off bias is about $-E_s/4$. The curves of

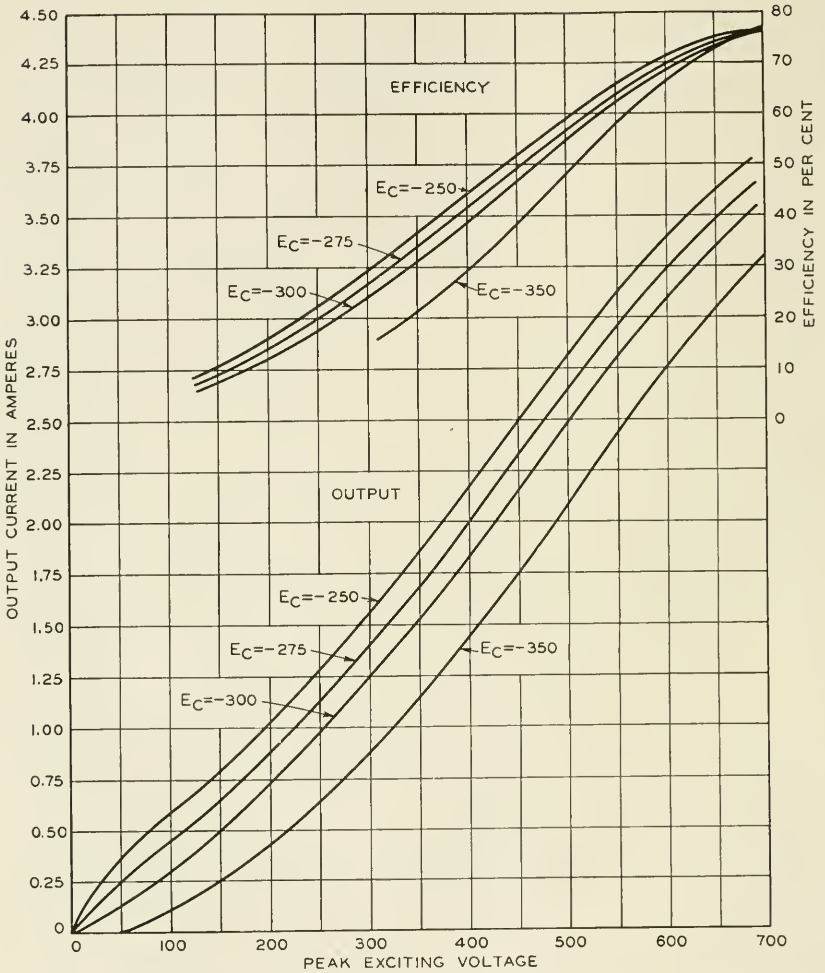


Fig. 13—Dynamic output and efficiency of a three-element tube. $E_b = 3000$ volts, $\mu = 10$, $Z_0 = 1500$ ohms.

Fig. 15 were taken with the screen voltage and output impedance constant at 400 volts and 1500 ohms respectively. It will be noted that these curves show the same general characteristics at the lower portions as the curves of the three-element tube, Fig. 13, but as predicted in the theoretical consideration, they tend to saturate more rapidly in the upper portion. The efficiencies appear to be about the same as obtainable from the three-element tube at the same outputs.

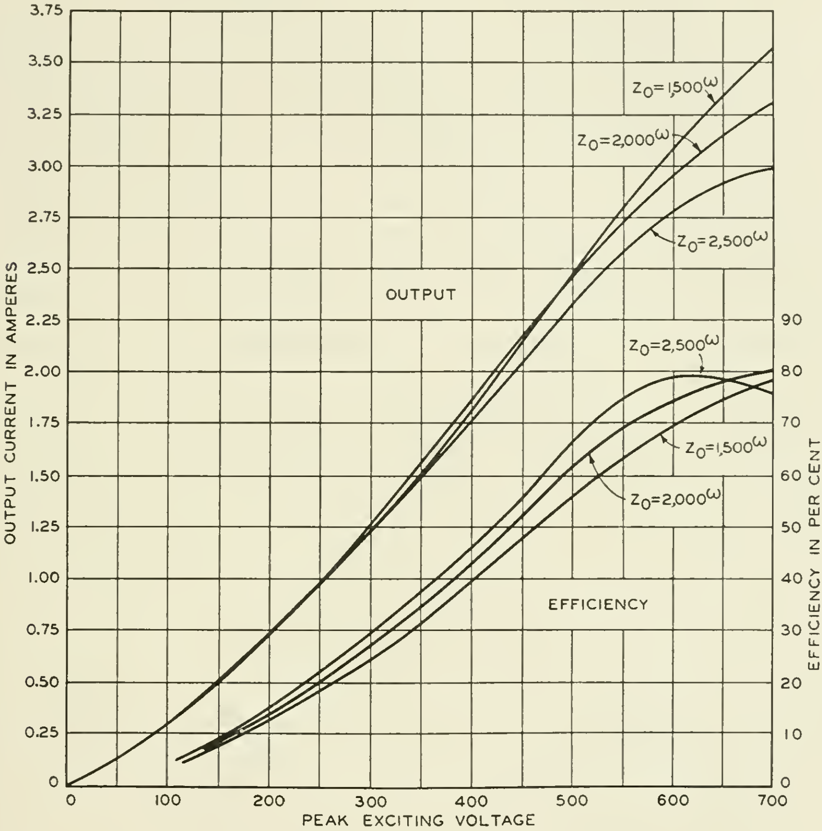


Fig. 14—Dynamic output and efficiency of a three-element tube. $E_b = 3000$ volts, $E_c = -300$ volts, $\mu = 10$.

Fig. 16 shows a comparison of the dynamic output curves at 1500 ohms impedance and bias at cut-off of the screen grid tube for three different values of screen voltage, and of the three-element tube. In addition the d-c. screen currents and grid currents are shown. The high μ effect of the screen grid tube is responsible for the steeper dynamic curve. The reason for the greater saturation effect at the upper portion of the curves for the screen grid tube is apparent upon comparing the d-c. screen and grid currents with the grid current for the three-element tube, since the plate is being robbed of the current taken by grid and screen in one case, and only of the current taken by the grid in the other. In Fig. 17, the effect of the output impedance on

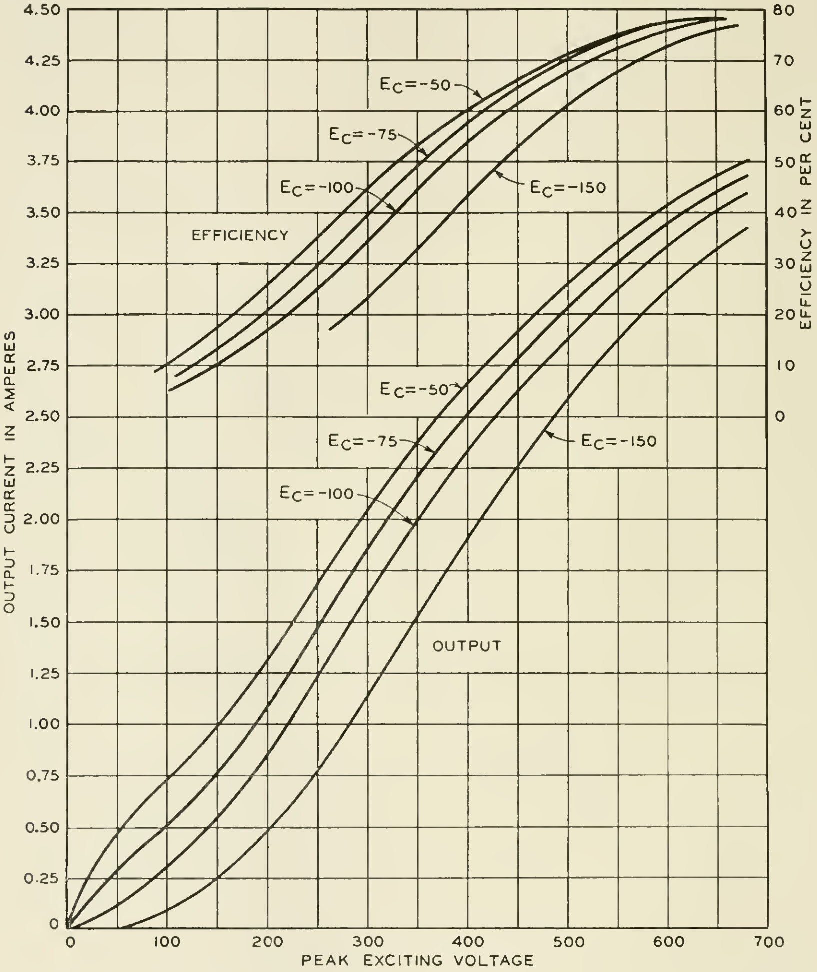


Fig. 15—Dynamic output and efficiency of a screen grid tube. $E_b = 3000$ volts, $E_s = 400$ volts, $Z_0 = 1500$ ohms.

the screen grid dynamic curves is shown. The effect is seen to be similar to that for the three-element tube, though somewhat more pronounced.

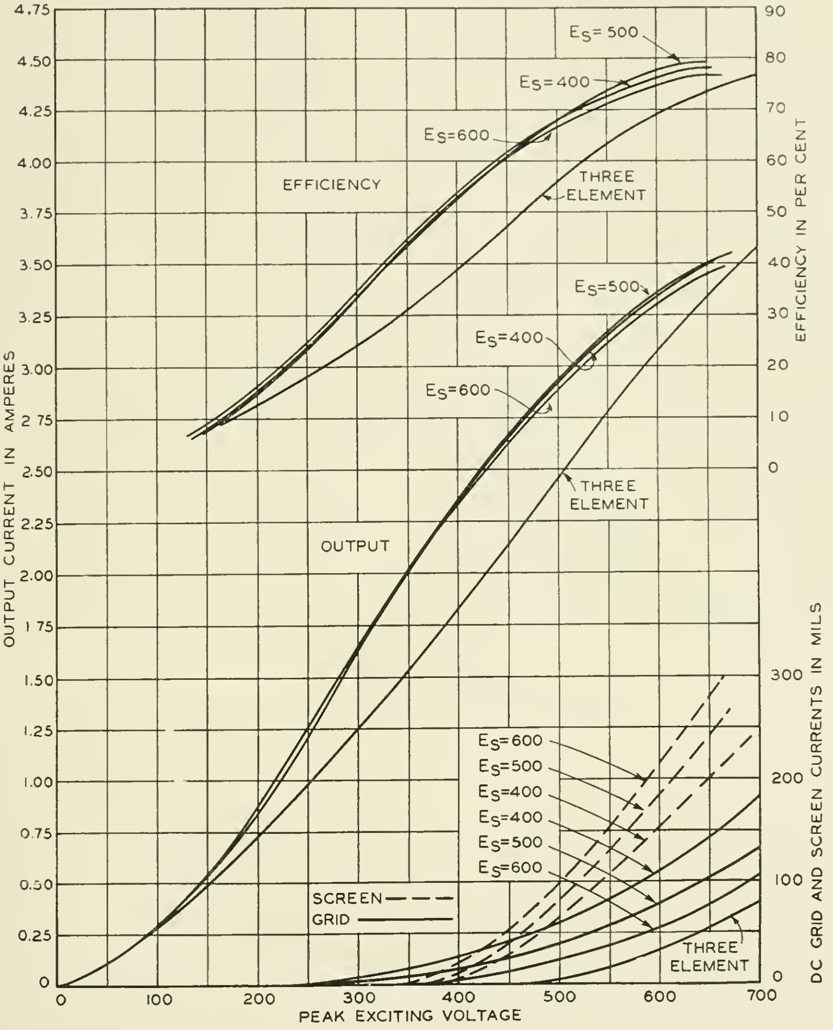


Fig. 16—Dynamic output and efficiency of a screen grid tube compared to that of a three-element tube. $E_b = 3000$ volts, E_c at cut-off $Z_0 = 1500$ ohms.

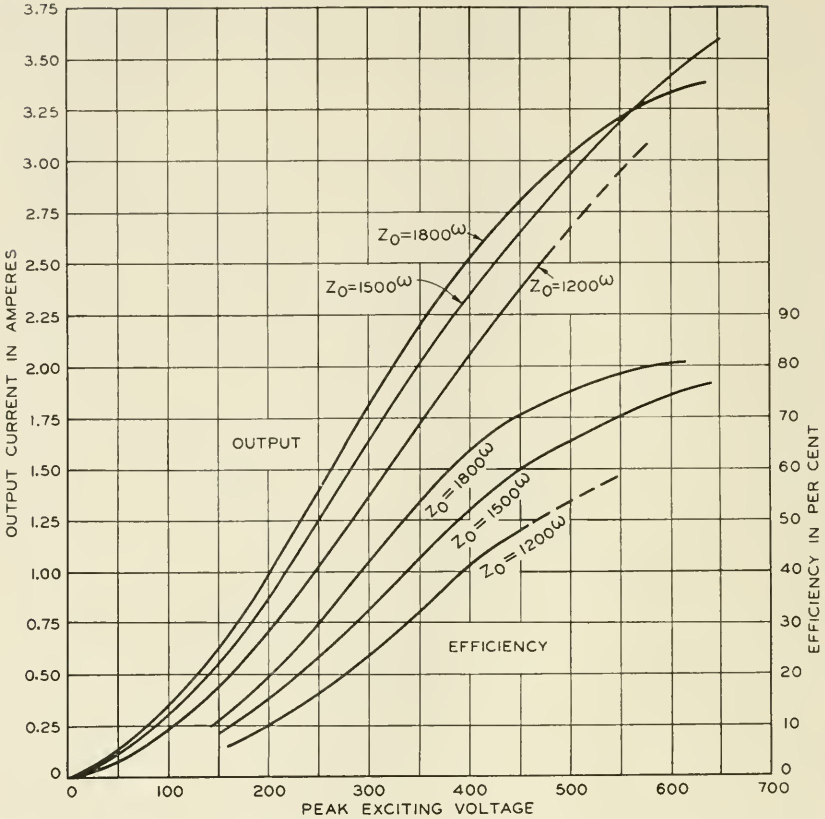


Fig. 17—Dynamic output and efficiency of a screen grid tube. $E_b = 3000$ volts, $E_c = -100$, $E_s = 400$ volts.

Conclusions

It has been shown from both theory and experiment that there is a marked difference between the dynamic output characteristics of Class B and Class C amplifiers as defined, particularly in the lower portion of the curves. This difference is the more pronounced the farther the grid bias voltage is moved from the cut-off point. In general, Class C operation is more efficient than Class B operation because the flow of plate current is limited to a smaller portion of the cycle which includes the portion in which the plate voltage is lowest.

It must be borne in mind that in a radio telephone transmitter where modulation is effected at low power level and the modulated carrier amplified, the dynamic characteristic of the output current will be the resultant of the dynamic characteristics of all of the intermediate stages from the modulating stage through the final amplifier stage. In certain cases distortion in one stage might be fairly well compensated for by suitably shaping the characteristic of the following stage through choosing proper values of bias voltage and output impedance.

It may be concluded from both theory and experiment that the screen grid tube functions in general similarly to the three-element tube and is capable of giving about the same output efficiency. It possesses the property of the high μ tube in giving greater output for a given exciting voltage in the portion of the dynamic characteristic where saturation has not become noticeable, but the dynamic characteristic shows saturation much sooner than that of the three-element tube. The screen grid tube also has the property that the output impedance has very little effect on the input circuit which will allow some shaping of the dynamic characteristic without affecting the driving power.

The fact should be mentioned that the so-called internal impedance, or plate impedance of the tube, such as might be measured on a bridge for small amplitudes of plate current swing, bears no very close relation to the impedance of the output circuit into which it should work in Class B or Class C operation. Reference to this plate impedance is quite misleading unless the part of the characteristic from which it is taken is given, since it varies greatly over the characteristic. For the screen grid tube mentioned the internal plate impedance is of the order of 100,000 ohms, whereas it was working with an output impedance of the order of 1500 ohms.

TABLE OF SYMBOLS

E_b = d-c. plate voltage
E_c = d-c. grid voltage
E_s = screen grid voltage (held constant)
e_p = instantaneous plate potential
e_g = instantaneous grid potential
e_{pm} = minimum plate potential
e_1 = instantaneous value of fundamental frequency component of alternating plate voltage
e_n = instantaneous value of n$^{\text{th}}$ harmonic component of alternating plate voltage
I_b = d-c. plate current
I_p = peak value of plate current
i_p = instantaneous value of plate current

i_0 = instantaneous value of alternating component of plate current

i_1 = instantaneous value of fundamental frequency component of alternating plate current

i_n = instantaneous value of nth harmonic component of alternating plate current

I_1 = peak value of fundamental frequency component of alternating plate current

I_n = peak value of nth harmonic component of alternating plate current

μ = amplification factor of tube = $\left[\dfrac{\Delta e_p}{\Delta e_g}\right]_{i_p \text{ constant}}$

K = I_1/I_p

W_0 = power output in watts

Z_0 = output impedance = R_0, a resistance at fundamental frequency

R_p = instantaneous internal resistance of the tube

r_p = differential plate resistance of tube at any point

$$= \left[\frac{\Delta e_p}{\Delta i_p}\right]_{e_g \text{ constant}}$$

The Time Factor in Telephone Transmission *

By O. B. BLACKWELL

Until comparatively recent years the telephone engineer gave little attention to transmission time in his problems. For all practical purposes he could assume that speech was transmitted instantly between the ends of telephone circuits. The rapid extension of the distances over which commercial telephony is given and the introduction of long telephone cables has changed the situation and has introduced time problems in telephone transmission which are of large technical interest and difficulty. As a result, time problems are receiving more consideration in the technical papers published in recent years on transmission. The accompanying bibliography lists a considerable number of such papers. There seems to be no paper, however, giving a general over-all picture of this subject. The present paper gives briefly such a picture.

THE time factor introduces five different types of problems in telephone transmission:

1. *A Slowing-Down of Telephone Communication.* In talking over long lengths of certain types of cable, the time interval between the formation of a sound by the speaker and its reception by the listener may become of sufficient magnitude to slow down conversation. This is not a serious matter with the types of circuits now used in the United States, even for the longest distances between points in this country. It does, however, become of considerable importance when we consider the joining together of long lengths of cable in this country and long lengths in Europe with possibly long lengths of intervening submarine cable.

2. *Delay Distortion.* Difference in the speed of transmission over a circuit of the different frequencies which make up speech. This may introduce peculiar distortions in speech which cause considerable interference.

3. *Echo Effects.* These arise from the fact that parts of the energy transmitted over a circuit may be reflected back from points of irregularity in it, particularly at the ends. Small amounts of the energy may wander back and forth over a circuit two or more times. While these echoes may affect both the talker and listener, they generally have the greatest effect on the talker who may have an uneasy feeling that the distant party wishes to break in on the conversation.

4. *Effects of Voice-Operated Devices.* To overcome echoes, and

---

* Presented at the South West District Meeting of the A. I. E. E., Kansas City, Mo., October 22–24, 1931.

under some conditions to hold circuits stable, it has become the practise of connecting into certain types of circuits, relay devices operated by the transmitted speech currents which render inoperative transmission in the opposite direction. In some cases delay in transmission may be an advantage in the operation of such devices. In other cases it may introduce serious difficulties. Conditions may be set up in which it is difficult for one party to interrupt the other. In other cases, portions of conversations may be locked out. If the voice-operated devices are not properly adjusted or if considerable noise is present, the devices may not function properly and speech mutilation may result.

5. *Fading.* In radio, the well-known phenomenon of fading is due to waves arriving at the receiver over different paths, the transmission times of which are such as to cause alternate strengthening and weakening of the received signal by alternate phase agreement and opposition. While this factor is mentioned here for completeness, it will not be discussed further as it is beyond the scope of this paper to discuss the problems introduced when there is more than a single path between the sending and receiving ends of a circuit. The present paper is limited to the conditions which hold where not more than one path is involved in the transmission in each direction.

SPEED OF TRANSMISSION

Before considering these problems in more detail it would be well to define what is meant by the speed of transmission over a circuit. There are several speeds which may have significance according to the problem involved.

Whenever a change in applied voltage is made at one end of a circuit, some evidence of this is transmitted over the circuit to the receiving end at the speed of light. In general, however, except in radio, no sufficient action to be of use is transmitted at this speed and it is largely of theoretical importance.

The speed which the engineer generally has in mind in thinking of line transmission is the speed at which the crests or the troughs of the waves pass along the line when a single-frequency potential is continuously applied at the sending end. This usually is referred to as the speed of phase transmission in the steady state. While this usually approximates the speed in which we are interested, it may in particular cases differ considerably from it. In fact, in certain types of artificial circuits the crests and troughs of the waves travel toward rather than away from the sending end.

This speed may best be explained as follows:

Since the speed of transmission in generally different for different parts of the frequency range, for simplicity a particular narrow frequency range, say between the frequencies N_1 and N_2, is considered. It is supposed that electrical filters are applied to the circuit limiting the frequencies over it to approximately this range. If then, a voltage having a frequency, say at midpoint of this narrow range, is applied to the circuit for a short interval and then removed, the speed at which the disturbance thus set up travels down the circuit is the speed in which we are interested. A spurt of energy of this type is evidently similar to that which takes place in carrier telegraph systems when a dot impulse is applied to the circuit. This speed can be looked at, therefore, as that of carrier telegraph signals so formed.

The speeds on this basis of a number of standard constructions which represent good engineering practise today are approximately as follows:

Type of circuit	Approximate speed in miles per second
Cable circuits loaded with 88-mh. coils at 3,000-ft. spacing	10,000
Cable circuits loaded with 44-mh. coils at 6,000-ft. spacing	20,000
Cable pairs of non-loaded 16 B. & S. gage	130,000
Non-loaded open-wire pairs	180,000
Radio	186,000

Causes of Time Lag in Transmission

A pair of wires of zero resistance in free space separated from all other conductors and without leakage would transmit electrical waves over it at the speed of light. It will be noted from the above table that non-loaded open wires transmit at a speed not differing widely from this. What retardation exists comes largely from the glass insulators which cause an increase in capacity and the resistance of the wires, which causes an effective increase in inductance.

In cable circuits there is still further retardation by the increase in the capacity between the wires because of the necessity of using a certain amount of solid dielectric and particularly from the increase in the inductance of the wires when loading coils are inserted in them to decrease attenuation.

In actual circuits there is still some further retardation by the apparatus which is necessarily inserted at the terminals and at intermediate points along the circuit. The figures given in the above table are for the bare circuits. The delays caused by apparatus will, in general, reduce these speeds from 10 to 25 per cent.

Slowing-Down of Telephone Conversation

Considering the first of the above factors, it is noted that so long as the speaker at one end of a telephone circuit continues to talk, the listener at the other end will hear the speech in proper time relation, independent of how much absolute delay there is in going from one end of the circuit to the other. However, when the speaker asks a question and waits for the answer, the slowing-down effect on his conversation will evidently be the time of transmission of his question to the distant end and the transmission back from the distant end of the answer. Considering the speed shown in the table, however, consideration may be taken of the fact that the non-loaded constructions, both open wire and cable, and the radio, are of such high speed that conversations could be carried on over them for the longest distance between places in the world without appreciable difficulty. This, however, is not the case for the loaded construction. Assume, for example, that a length of 4,000 miles would cover the wire line distance between any two points in this country. For the slowest construction noted, an interval of 0.8 second is required for transmission to the distant end and return. While it is possible to carry on conversation over a circuit with this delay, it is larger than is considered desirable. The faster of the loaded constructions shown would give a delay over circuits of this length which represents somewhere about the limit of what, at the present time, is considered satisfactory. Incidentally, the slowest of the constructions shown, for this and other reasons, is not proposed for use except for comparatively short distances.

Communication engineers must look forward to the time when the longest cable distances in North America are connected, in some cases by submarine cable, to long lengths of cable in Europe. With this as an ultimate objective, this matter of the direct effect of delay on conversation has become of considerable interest.

An appreciation of the transmission time on long telephone circuits may perhaps be gained by considering the distance required to produce an equivalent delay of sound waves traveling in air. For example, it takes about as long for a radio wave to travel half way around the earth at the equator as it does for a sound wave to travel from one speaker to another when the distance separating them is about 75 ft. Incidentally, the time required for a radio wave to travel from the earth to the planet Mars would be from about 3 to 20 minutes, assuming that it got there at all. Evidently if we have neighbors on Mars we can never hope to carry on conversation with them.

Telephone engineers have devoted considerable attention to the effects of delay on the telephone users in an effort to determine how far the electrical waves should be permitted to travel over different types of circuits. Since the present constructions do not offer any particular difficulties, for the distances now in use, a look into the future has been taken by means of artificial delay circuits. One method has been to loop back and forth loaded conductors in a cable until the desired delay was obtained Another method has been to record the

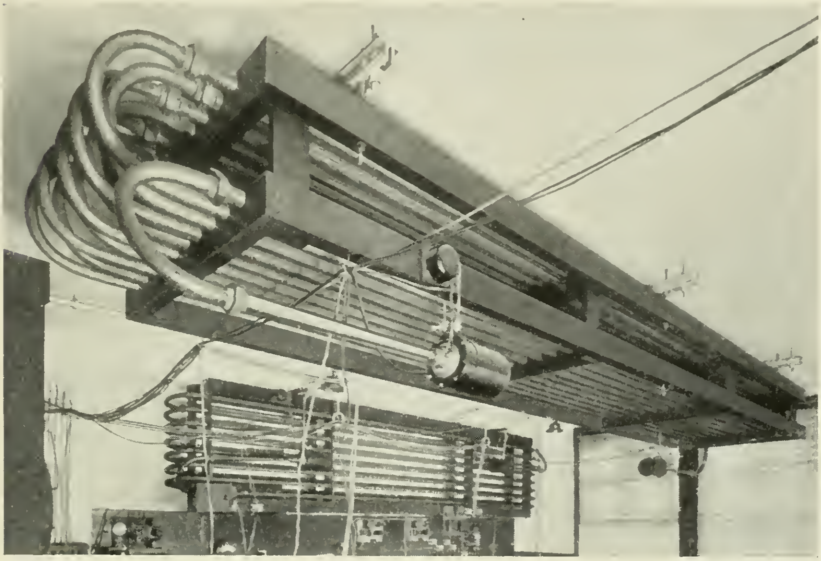


Fig. 1—Acoustic delay circuits.

talkers' waves on a phonograph and pick up the impressions with a second needle, displaced the desired amount from the first so as to introduce delay into the conversation.

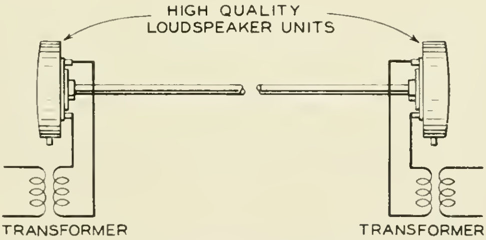


Fig. 2—Acoustical delay circuit showing two high-quality loudspeakers with one connected to each end of tube forming delay circuit.

Considerable use has also been made of pipes or "acoustic" delay circuits. Fig. 1 shows an illustration of a brass pipe delay circuit used in experimental work. Fig. 2 shows the circuit in schematic form. In addition to the pipe, which is looped back and forth to conserve space, the circuit involves telephone receivers at the two ends. The one at the sending end converts electrical energy into sound which is transmitted through the pipe and the one at the receiving end converts the sound waves back into electrical energy. The pipes are quite suitable because they have approximately the same delay at all frequencies. Various devices are required to reduce the reflections which occur at the junction of the pipe and the receiver and to equalize the attenuation.

Using devices of this nature, experimenters have found it possible to talk fairly conveniently over circuits representing time intervals as great as 0.7 second in each direction. So great delays would be considered undesirable for commercial use, however. Delays of about a third of this, in general, are considered about the maximum which is satisfactory.

Delay Distortion

In designing circuits which are electrically long, care must be exercised to insure that the transmission times for all frequencies in the transmission range are sufficiently alike to avoid objectionable transient phenomena. These effects may occur in one-way circuits as well as in two-way circuits and are not related to echo effects.

The appearance of these transients to the listener depends on whether the excess delay is at low frequencies or at high frequencies. It is rather difficult to describe the characteristic sound of a circuit with low-frequency delay. A high-frequency delay, if it is in an extreme form, sounds as though a high-pitched reed, such as a harmonica reed, was being plucked whenever there is a sudden transition in the voice sounds being transmitted over the circuit.

The characteristic effects of transients are conveniently described by the aid of oscillograms of spurts of alternating current taken before and after being sent over circuits having various delay characteristics. To begin with, it must be recalled that when any wave shape is applied to a circuit, the transmitted wave in the circuit can be expressed as the sum of the series of sinusoidal waves whose frequencies range from very low to very high values.

In the case where a sinusoidal wave of frequency F is suddenly applied to the sending end of the line, the effect may therefore be explained as due to an infinity of sinusoidal waves so proportioned and phased as to add up to zero, up to the instant of application of the wave, and to equal the steady-state value of the wave at that instant. Of these waves, the most important have frequencies close to F. They are propagated over the line individually with a velocity corresponding to the frequency. If the velocity of the line is the same for all frequencies, they will evidently add up to give the same over-all wave shape at the receiving end of the line as at the sending end.

If, on the other hand, the velocity is not the same for all frequencies, there will be more or less distortion and transients in establishing the wave, though ultimately the pure wave of frequency F will be established. Several oscillograms will be shown to indicate transient effects which are experienced under various conditions.

Fig. 3 is an oscillogram showing a spurt of 1,600-cycle current as applied to and received from a loaded circuit having fairly large delays in the upper part of the transmitted range compared to the delay at lower frequencies. Remembering the nature of the oscillations at the beginning and end of the applied spurt, it will be observed that the current at the receiving end consists at first of a fairly low frequency which builds up in frequency and magnitude to the steady-state value. At the end of the spurt the same transient is experienced, but in this case the higher frequency currents which have been delayed in the line are at the tail end of the train.

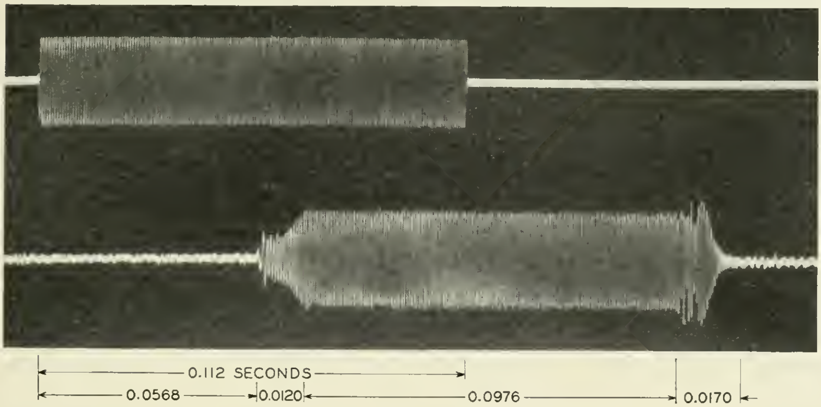


Fig. 3—Transients in 522 miles of medium heavy loaded repeatered circuit. Upper trace—transmitted 1600-cycle wave. Lower trace—received wave.

Fig. 4 shows a 200-cycle current with many harmonics of higher degree transmitted over a circuit having large delay at low frequencies. It will be noted that these high or harmonic frequencies are received in advance of the 200-cycle wave. This is because the 200-cycle wave is subject to appreciable delay while the higher frequencies are not. This circuit, while actually made up of artificial networks, had characteristics similar to certain types of long cable circuits for the lower part of the telephone frequency range.

Fig. 5 shows transient effects of a 600-mile composited 19-gage H-174 side circuit. The zero lines of the three curves of this figure show slight effects of crosstalk and other interference. These effects are not, however, of sufficient magnitude to interfere with the general appearance of the signals. The upper line shows the applied 1,000-cycle current. The next line shows the current as it was received at the end of the line. The transients which are produced at the beginning and ending of the signals are evident.

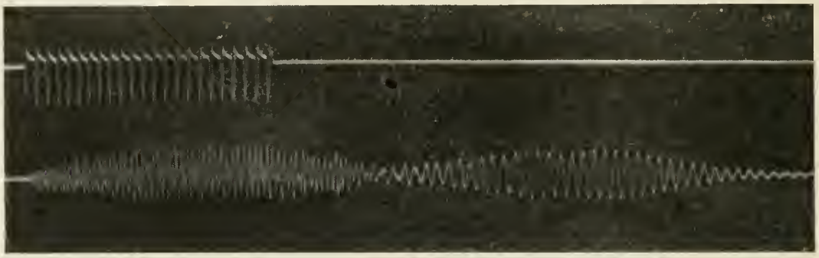


Fig. 4—Transients in 174 sections of high-pass filter (cutoff frequency 107 cycles). Upper trace—transmitted 200-cycle wave with harmonics obtained from an overloaded amplifier. Lower trace—received wave.

By the insertion of proper networks it is possible to correct for distortions of this kind. In the last line there is shown the received current when a delay correcting network of this kind is applied in series with the circuit. It should be noted that while the delay in the reception of the signal is somewhat increased (as shown by the displacement

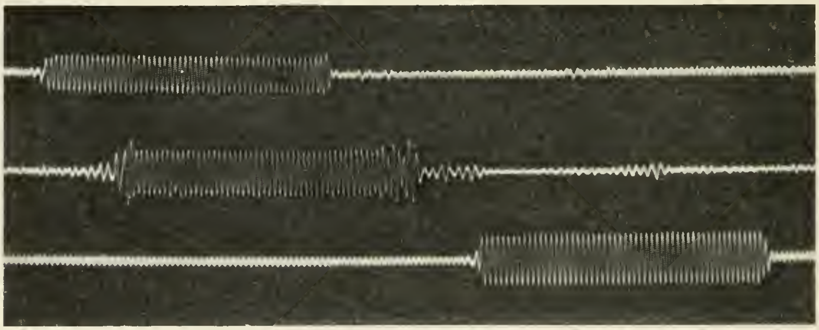


Fig. 5—Correction of transients in 600 miles of medium heavy loaded cable circuit through use of delay correcting network. Upper trace—transmitted 1000-cycle wave. Middle trace—wave received from line and applied to delay correcting network. Lower trace—wave received from delay correcting network.

of the whole signal farther to the right) the transients at the beginning and stopping of the signal are very much reduced.

ECHOES

In designing telephone circuits which are electrically long, an important problem is presented by the necessity of avoiding echo effects. These are caused by reflection of electrical energy at points of discontinuity in the circuit and are very similar to echoes of sound waves in an auditorium. The reflected waves are usually considered as echoes when there is an appreciable delay with respect to direct transmission. Some of the reflected waves return to the receiver of the talker's telephone so that if the effects are severe he may hear an echo of his own words. Other reflected waves enter the receiver of the listener's telephone and, if severe, cause the listener to hear an echo following the directly received transmission.

Reflections of voice waves occur in all practical telephone circuits. It is only in telephone circuits of such length as to require a number of repeaters, however, that echo effects become serious. The fact that the circuits are electrically long makes the time lag of the echoes appreciable. At the same time, the telephone repeaters overcome the high attenuation in these long circuits and consequently make the echoes louder. The seriousness of the effect is a function of both the time lag and the volume of the echo relative to the direct transmission, becoming greater when these are increased.

In telephone circuits the most important points of discontinuity are usually the two ends of the circuit. In a four-wire telephone circuit these are the only points of discontinuity.

Fig. 6 shows a schematic diagram of a four-wire telephone circuit and a schematic representation of the direct transmission over the circuit, together with the various talker and listener echoes which are set up. The rectangles at the extreme right and left are intended to represent the telephone sets used by two subscribers at the west and east terminals of the circuit. The rectangles marked *N* represent electrical networks which simulate or balance more or less perfectly the impedance of the telephone sets. In the four-wire circuit the rectangles with arrows represent one-way repeaters or amplifiers. At each terminal the two separate one-way circuits comprising the four-wire circuits are joined together by means of the familiar balanced transformers. When the subscriber at *W* talks, the transmission passes to *E* over the upper path in the four-wire circuit. This is indicated by the heavy line labeled "Direct Transmission" in part *b* of the figure. When subscriber *E* talks, transmission passes over the lower path in a similar manner.

Considering part *b* of the figure, it will be noted that when direct transmission is received at the east end of the circuit, a portion of the current passes to the opposite side of the four-wire circuit and is transmitted to the subscriber at the west end as a talker echo. Similarly, a portion of this talker echo is transmitted over the upper part of the circuit to the listener at the east end of the circuit as a listener echo. Successive talker and listener echoes follow this, as indicated in the diagram. If the networks at the two ends of the circuit can be made to simulate accurately the subscriber circuits, none of these echoes will exist. A high degree of simulation, however, is impracticable in an economical telephone plant under usual conditions. In two-wire circuits with many repeaters the echo paths may become very complicated.

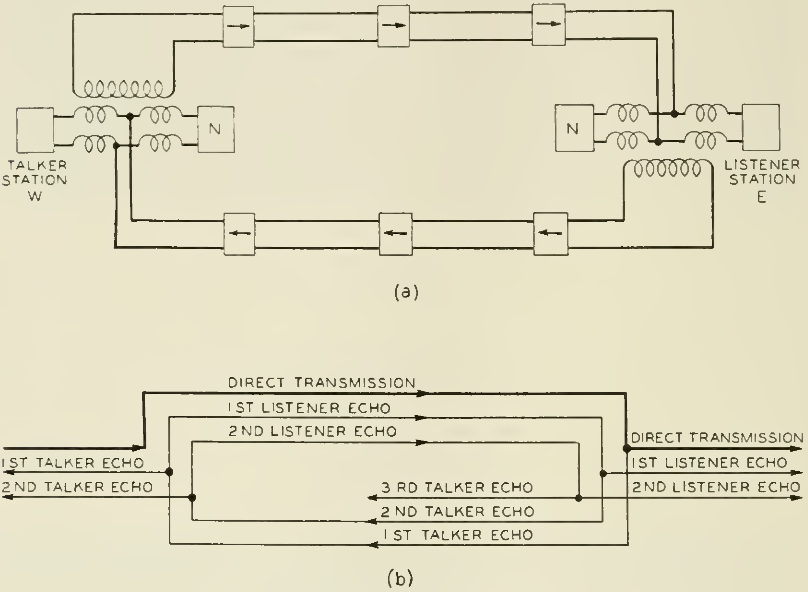


Fig. 6—Echoes in four-wire circuit.

An interesting case of "echoes" is that which may be produced when two radio stations are sending out the same program at the same wavelength. The program as received by one of these stations over wire circuits is, of course, slightly delayed with respect to a station nearer the source of the program. It is possible, then, for a receiving set properly located to receive both of these stations, in which case if the time difference is sufficient one of them will sound like an echo of the other. In this case there may be the added peculiarity that if the weaker station is the one at which the program is received first it will appear that the "echo" is in advance of rather than following the sound which appears to cause it.

Time Effects with Voice-Operated Devices

Switching devices operated by the voice currents themselves are frequently introduced into long telephone circuits. In general, the effect of such devices is to render inoperative transmission in the direction opposite to that of the speech waves which are going over a circuit at the particular instant. The first use of any considerable importance to which such devices were put was in connection with long circuits for the purpose of preventing the building-up of undesirable echoes. More recently, however, long radio telephone circuits have come into use. These circuits may vary rapidly in transmission effectiveness. If these circuits are arranged to be operative in both directions at a time it would be very difficult to prevent their becoming unstable and possibly setting up oscillations. For this reason such circuits are frequently operated with switching arrangements such that the circuits leading to both transmitting stations are normally disabled and rendered inoperative. When a subscriber speaks at either end, therefore, the voice currents must operate switching devices which restore the circuits leading to his transmitting station. Incidentally, this must render inoperative the receiving circuit at the same time.

A very interesting application of time delay has been made in connection with radio systems so operated. In this arrangement the voice currents when they reach the disabling point are passed through an artificial line in which a desired amount of delay is incorporated. Just before entering this line a fraction of the energy is taken, rectified, and made to operate the switching mechanism for restoring the circuit to operating condition. This switching is so arranged as to be completed by the time that the voice currents have passed through the artificial delay circuit and are ready to proceed down the line. If it were not for this arrangement a small part of the speech currents might be dissipated during the interval while the switching mechanisms were operating.

Fig. 7 is similar to Fig. 6 noted above with the exception that the application of an echo suppressor is shown.

When the subscriber at the left of the drawing begins to talk, the waves set up in his telephone are transmitted over the upper part of the circuit. Upon reaching the input of the echo suppressor, a small part of the energy is diverted to short circuit the lower branch of the circuit, as indicated. Meanwhile, the main transmission passes on to the subscriber at the right. Echoes which return in the lower part of the circuit are blocked as indicated. After the talker has ceased speaking, the device remains operative for a time equal to the delay of the echo as measured from the input of the device to the disabling point plus an additional time to take care of echoes in the circuit between the four-wire terminal and the subscriber.

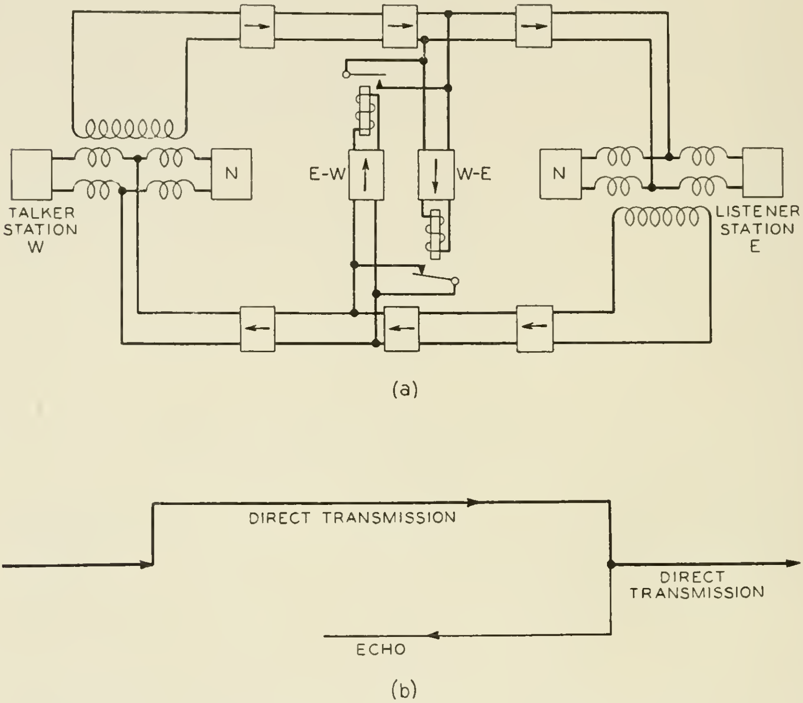


Fig. 7—Echo suppressor cutting off echo in four-wire circuit.

When the suppressor releases, the circuit is again free to transmit in either direction. When the right-hand subscriber talks, the action is similar except that the other half of the echo suppressor operates.

In practical use the echo suppressors are so carefully controlled that telephone users are generally unable to tell whether a suppressor is on the circuit or not. This is due to the short delays and careful adjustments of the time functions of the device. It is possible, however, if the delays are longer and the adjustments are made with less care, to introduce two types of difficulty. If one subscriber talks fairly steadily he may hold the suppressor operative so continuously that it is difficult for the other party to break in on the conversation if he so desires. If one subscriber started to reply almost simultaneously with the termination of the other's speech, part of his reply might be blocked at the echo suppressor along with the last of the echo.

Further difficulties arise if two circuits, each containing an echo suppressor of this kind, are switched together in tandem. In this case it would be possible for the subscribers to completely block each other's speech if they started talking simultaneously.

In the case of radio circuits operated as noted above, difficulties are introduced somewhat similar to that of two echo suppressors in tandem.

The above will sufficiently suggest the types of difficulties which arise from delay in connection with voice-operated devices. Certain of the papers in the accompanying bibliography consider these problems in more detail.

In preparing the accompanying bibliography, no attempt has been made to make it complete. It is believed, however, to contain most of the important publications on the subject.

BIBLIOGRAPHY

General

"Telephone Transmission over Long Cable Circuits," A. B. Clark. *Bell System Technical Journal*, vol. 2, no. 1, Jan. 1923, p. 67–94; A. I. E. E. Trans., vol. 42, Feb. 1923, p. 86–96; *Electrical Communication*, vol. 1, no. 3, Feb. 1923, p. 26–40.

"Development and Application of Loading for Telephone Circuits," Thomas Shaw and William Fondiller. *A. I. E. E. Trans.*, vol. 45, Feb. 1926, p. 268–292; *Bell System Technical Journal*, vol. 5, no. 2, April 1926, p. 221–281; *Electrical Communication*, vol. 4, no. 4, April 1926, p. 258–276, and vol. 5, no. 1, July 1926, p. 38–53.

"Extending the Range of Loaded Lines by Echo Suppression and Phase Equalization," K. Küpfmüller. *Elektrische Nachrichten-Technik*, vol. 3, no. 3, Mar. 1926, p. 82–88.

"Transmission Characteristics of Long Distance Telephone Cable Connections," K. Küpfmüller. *Europäischer Fernsprechdienst*, no. 5, July 1927, p. 19–33.

"Telephone Communication System of the United States," Bancroft Gherardi and F. B. Jewett. *Bell System Technical Journal*, vol. 9, no. 1, Jan. 1930, p. 1–100.

"Recent Developments in Toll Telephone Service," W. H. Harrison. *A. I. E. E. Trans.*, vol. 49, no. 1, Jan. 1930, p. 166–176.

"Some Recent Developments in Long Distance Cables in the United States of America," A. B. Clark. *Bell System Technical Journal*, vol. 9, no. 3, July 1930, p. 487–492.

"Transmission Characteristics of Open-Wire Telephone Lines," E. I. Green. *Bell System Technical Journal*, vol. 9, no. 4, Oct. 1930, p. 730–759.

"Range and Economy of the Toll Cable Lines of System II of the C. C. I.," K. Höpfner and F. Lüschen. *Europäischer Fernsprechdienst*, no. 21, Jan. 1931, p. 3–16.

Delay Distortion

"Theory of the Transient Oscillations of Electrical Networks and Transmission Systems," J. R. Carson. *A. I. E. E. Trans.*, vol. 38, pt. 1, 1919, p. 345–427.

"Transient Phenomena in Wave Filters," K. Küpfmüller. *Elektrische Nachrichten-Technik*, vol. 1, no. 5, Nov. 1924, p. 141–152.

"Measurement of Phase Distortion," Harry Nyquist and Smart Brand. *Bell System Technical Journal*, vol. 9, no. 3, July 1930, p. 522–549.

"Building-Up of Sinusoidal Currents in Long Periodically Loaded Lines," J. R. Carson. *Bell System Technical Journal*, vol. 3, no. 4, Oct. 1924, p. 558–566.

"Attenuation of Band-Pass Filters in the Transmission Range," H. F. Mayer. *Elektrische Nachrichten-Technik*, vol. 2, no. 10, Oct. 1925, p. 335–338.

"Phase Distortion and Phase Distortion Correction," S. P. Mead. *Bell System Technical Journal*, vol. 7, no. 2, April 1928, p. 195–224.

"Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks," O. J. Zobel. *Bell System Technical Journal*, vol. 7, no. 3, July 1928, p. 438–534.

"Effects of Phase Distortion on Telephone Quality," J. C. Steinberg. *Bell System Technical Journal*, vol. 9, no. 3, July 1930, p. 550–566.

"Phase Distortion in Telephone Apparatus," C. E. Lane. *Bell System Technical Journal*, vol. 9, no. 3, July 1930, p. 493–521.

Echoes

"Irregularities in Loaded Telephone Circuits," George Crisson. *Bell System Technical Journal*, vol. 4, no. 4, Oct. 1925, p. 561–585; *Electrical Communication*, vol. 4, no. 2, Oct. 1925, p. 98–110.

"Limitation of the Gain of Two-Way Telephone Repeaters by Impedance Irregularities," George Crisson. *Bell System Technical Journal*, vol. 4, no. 1, Jan. 1925, p. 15–25.

"Connecting-Together of Four-Wire Lines," K. Höpfner. *Telegraphen- und Fernsprech-Technik*, vol. 16, no. 6, June 1927, p. 166–169.

"Measurement of Echo Attenuation," R. Feldtkeller and H. Jacoby. *Telegraphen- und Fernsprech-Technik*, vol. 17, no. 3, March 1928, p. 61–71.

Voice-Operated Devices

"Echo Suppressors for Long Telephone Circuits," A. B. Clark and R. C. Mathes. *A. I. E. E. Trans.*, vol. 44, April 1925, p. 481–489; *Electrical Communication*, vol. 4, no. 1, July 1925, p. 40–50.

"Echo Suppressors for Telephone Lines," H. F. Mayer and H. Nottebrock. *Telegraphen- und Fernsprech-Technik*, vol. 15, no. 11, Nov. 1926, p. 353–355.

"New York-London Telephone Circuit," S. B. Wright and H. C. Silent. *Bell System Technical Journal*, vol. 6, no. 4, Oct. 1927, p. 736–749.

"Transatlantic Telephony. The Post Office Differential Voice-Operated Anti-Singing Equipment," C. A. Beer and G. T. Evans. *P. O. E. E. Journal*, vol. 20, no. 1, April 1927, p. 65–72.

"Bridge for Measuring Small Time Intervals," Joseph Herman. *Bell System Technical Journal*, vol. 7, no. 2, April 1928, p. 343–349.

"New Standard Echo Suppressor," W. F. Marriage, P. R. Thomas, and K. G. Hodgson. *Electrical Communication*, vol. 9, no. 3, Jan. 1931, p. 196–202.

Related Transmission

"Transmission of Pictures over Telephone Lines," H. E. Ives, J. W. Horton, R. D. Parker, and A. B. Clark. *Bell System Technical Journal*, vol. 4, no. 2, April 1925, p. 187–214.

"Some Studies in Radio Broadcast Transmission," Ralph Bown, DeL. K. Martin, and R. K. Potter. *Bell System Technical Journal*, vol. 5, no. 1, Jan. 1926, p. 143–213.

"Wire Transmission System for Television," D. K. Gannett and E. I. Green. *Bell System Technical Journal*, vol. 6, no. 4, Oct. 1927, p. 616–632.

"Some Developments in Common Frequency Broadcasting," G. D. Gillett. *Proceedings of the Institute of Radio Engineers*, vol. 19, no. 8, Aug. 1931, p. 1347–1369.

Constant Frequency Oscillators *

By F. B. LLEWELLYN

Summary—The manner in which the frequency of vacuum tube oscillators depends upon the operating voltages is discussed. The theory of the dependence is derived and is shown to indicate methods of causing the frequency to be independent of the operating voltages. These methods are applied in detail to the more commonly used oscillator circuits.

Experimental data are cited which show the degree of frequency stability which may be expected as a result of application of the methods outlined in the theory, and also show that the best adjustment is in substantial agreement with that predicted by theory. With a carefully built and adjusted oscillator the effects of normal variations in the operating voltages are negligible in comparison with the effects of temperature variations resulting from the changed operating currents. Methods of preventing these latter effects are not discussed in the present paper.

The appendix contains an analysis of the conditions under which the performance of an oscillator may be represented by the use of linear circuit equations.

IN recent years the commercial requirements of vacuum tube oscillators have grown more rigid. The tremendous increase in the number of radio broadcast stations with the consequent narrowing of frequency band available to each, the analogous demands by the carrier telephone, and the tendency toward higher frequencies where a small percentage frequency change defeats the universal effort to secure better quality, all have united in creating a need for very constant frequencies. This need has led to a study of methods for holding the frequency constant. The most notable of these is the piezo-electric crystal. However, it has been known for some time that certain oscillator circuits have the inherent property of maintaining their frequency quite constant even though not crystal controlled. Some of these circuits have the additional advantage of combining constant frequency at a given wave-length with the ability to maintain this constancy at other wave-lengths, thus giving a range of available frequencies, any one of which may be depended upon to stay constant.

The elements which cause the frequency of oscillators which are not crystal controlled to vary are such things as vibration, changing temperature, fluctuating voltage, and changing load. Vibration and temperature affect primarily the inductance and capacity in the circuit which naturally causes the frequency to change. Fluctuating voltages change the tube resistance, which in turn affects the frequency.

---

* Presented at Sixth Annual Convention of I. R. E., Chicago, Illinois, June 4–6, 1931. Published in *Proc. I. R. E.*, Dec., 1931.

Changing loads also change the frequency, since they take the form of variable resistance and reactance.

Considerable work has been done by various individuals to make the inductance and capacity of standard apparatus as free from the effects of vibration and temperature changes as possible. Work of that kind is not discussed here. Variable voltages occur in practically all installations and changing load impedance in many. These two things are the actual cause of the larger part of frequency variation in many installations.

Several vacuum tube circuits have been devised to surmount these difficulties. They may be divided roughly into two groups: first, those in which the attempt has been to minimize the change of frequency with battery voltage, and second, those in which the attempt has been to prevent the change. In the first group we have two types: first, circuits in which the effective resistance has been reduced to as low a figure as possible, and second, those in which a high impedance has been inserted between the tube and the tuned circuit in order to reduce the relative effect of changes in the tube. Considerable success has attended the efforts of a number of engineers led by J. W. Horton in these directions. More recently, circuits of the first type have been applied to the production of relatively high frequencies.[1]

The second group in which the attempt has been to prevent the frequency change developed from the work of Messrs. J. F. Farrington and C. F. P. Rose. They found that a certain critical value of an impedance between the vacuum tube and the tuned circuit apparently produced a constant frequency over a limited range when the battery voltages were varied. They experimented with various forms of networks for this stabilizing impedance and developed several in which the output power was not reduced by stabilization.

THEORY

The writer attacked the problem from a theoretical standpoint and showed that in certain cases the mathematical procedure indicates means of making the oscillator frequency independent not only of a variable load resistance, but also of the battery voltages. The purpose of this paper is to develop the general theory and application of these circuits and to show how several circuits in particular may be made to produce practically constant frequency with customary variations of voltage and load resistance. The relations necessary to maintain the frequency constant at any given setting when it is desirable that the oscillator be operative over a range of frequencies are also indicated.

---

[1] Ross Gunn, "A new frequency stabilized oscillator system," *Proc. I. R. E.*, **18,** September, 1930.

Before proceeding with a detailed description of the various specific embodiments necessary to secure independence of frequency and battery voltage, it will be well to lay down the physical conditions upon which the frequency of any vacuum tube oscillator depends.

In general, all such oscillators consist of or may be resolved into, a tuned electrical circuit or network to which is attached a vacuum tube. Irrespective of any particular circuit, the frequency of the oscillator is completely determined by the following quantities, the designations used here being uniformly employed throughout the subsequent analysis:

L, the self-inductance in the network
M, the mutual inductance in the network
C, the capacity in the network
R, the resistance in the network
r_p, the plate resistance of the vacuum tube
r_g, the grid resistance of the vacuum tube
μ, the amplification factor of the vacuum tube

Of these quantities, L, C, and M require little comment. They are merely symbolic of the elements of the electrical network. The quantity C includes the interelectrode capacities of the tube when they become of consequence. These three quantities are assumed to be constant, an assumption which has been found very reasonable in practice. The quantity R represents the resistance in the network. For the purpose of this discussion the oscillator is assumed to deliver only a small amount of power, being used most often in such a manner as to supply voltage to the grid of an amplifier tube. Consequently, the electrical network external to the vacuum tube may, and should, be constructed in such a manner as to include a minimum amount of resistance. Under these conditions the losses in the circuit have been found to be practically all the result of the internal resistances, r_p and r_g of the vacuum tube.

These two quantities, r_p and r_g, are very important, being principally responsible for changes in condition of the circuit as a whole. It should be realized that r_g has the same relation to the static values of grid current and potential that r_p has to the plate current and potential. The effect of varying the applied potential of the grid or plate, or of changing the filament current is directly to cause r_p and r_g to vary, usually in opposite directions. Further, when amplitude of oscillation varies, for which variation of battery voltages (grid, plate, and filament) are again principally responsible, both r_g and r_p vary.[2]

The quantity μ is the amplification factor and is used with its usual significance. It varies with battery potential but this variation is ordinarily very small, though not to be neglected.

It eventuates, from the above considerations, that if the reactive elements of the frequency determining circuit are constant, a permissible assumption, the frequency may be stabilized if adequate account is taken of changes in battery voltages and load resistances. This it is the purpose of the present paper to discuss.

HARTLEY OSCILLATOR

Consider first the form of the Hartley oscillator shown in schematic form without indicating any special method of introducing the batteries, in Fig. 2. Figure 1 shows the circuit equivalent of several of the oscillators in the following figures when the impedances are represented in generalized form, and therefore will be employed for an analysis of the conditions necessary to secure independence of frequency and battery or applied voltages, and the results of this analysis will then be interpreted in detail in terms of the special circuit of Fig. 2. In Fig. 1 the impedances, Z_4 and Z_5, are inserted for the purpose of effecting the independence of frequency and battery voltages, and the values which they should have in order to accomplish this result are found by the following analysis:

From Fig. 1 we have the circuit equations when the assumed current conditions are as shown by the arrows:

$$\begin{aligned}
\mu e &= I_1(r_p + Z_1 + Z_5) + I_2(Z_1 + Z_m) - I_3 Z_m, \\
0 &= I_1(Z_1 + Z_m) + I_2 Z_0 - I_3(Z_2 + Z_m), \\
0 &= -\, I_1 Z_m \qquad - I_2(Z_2 + Z_m) + I_3(r_g + Z_2 + Z_4), \\
e &= I_3 r_g.
\end{aligned} \tag{1}$$

These equations are expressions of Kirchkoff's Law regarding the sum of the potentials in a closed mesh. The equations (1) effectively comprise only three simultaneous equations because the network has only three meshes.

In the above equation Z_0 is symbolic of the series impedance of the tuned circuit. Using the symbolism of Fig. 1,

$$Z_0 = Z_1 + Z_2 + Z_3 + 2Z_m. \tag{2}$$

---

[2] The appendix to this paper contains a further discussion of the significance of r_p and r_g together with an analysis of the conditions under which oscillator networks may be treated by the use of linear circuit equations as is done in the following analysis.

Equations (1) may be rewritten in determinant form as follows:

$$\begin{vmatrix} (r_p + Z_1 + Z_5) & +(Z_1 + Z_m) & -(Z_m + \mu r_g) \\ +(Z_1 + Z_m) & Z_0 & -(Z_2 + Z_m) \\ -Z_m & -(Z_2 + Z_m) & (r_g + Z_2 + Z_4) \end{vmatrix} = 0. \qquad (3)$$

This determinant form of (1) follows immediately from reducing (1) to three equations.

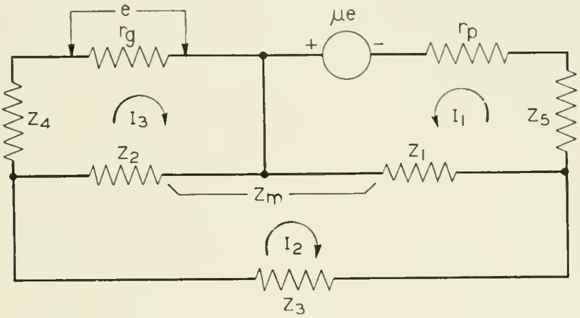


Fig. 1—Equivalent circuit network of Hartley or Colpitts-type oscillator.

In accordance with the theory of the operation of oscillators, discussed in the appendix, both the conditions necessary for oscillation to exist and the frequency of oscillation may be found from (3). That is:

$$\begin{aligned}(r_p + Z_1 + Z_5)Z_0(r_g + Z_2 + Z_4) &+ (Z_1 + Z_m)(Z_2 + Z_m)(\mu r_g + 2Z_m) \\ = Z_0 Z_m(\mu r_g + Z_m) &+ (Z_1 + Z_m)^2(r_g + Z_2 + Z_4) + (Z_2 + Z_m)^2 \\ &(r_p + Z_1 + Z_5). \qquad (4)\end{aligned}$$

The next step is to express each of the generalized Z's in the equivalent form of $(R + iX)$ where i stands for the imaginary quantity, $\sqrt{-1}$, and both R and X are real, representing, respectively, resistance and reactance. A great simplification results when it is recalled that the circuits external to the vacuum tube are assumed to have very little resistance, and that practically all of the losses in the network are caused by the tube resistances, r_g and r_p, so that these two are the only resistances which need be retained in the analysis. With this understanding, (4) becomes:

$$\begin{aligned}&[r_p+i(X_1+X_5)]iX_0[r_g+i(X_2+X_4)]-(X_1+X_m)(X_2+X_m)(\mu r_g+2iX_m) \\ &\quad = -X_0X_m(\mu r_g+iX_m)-(X_1+X_m)^2[r_g+i(X_2+X_4)] \\ &\quad -(X_2+X_m)^2[r_p+i(X_1+X_5)]. \qquad (5)\end{aligned}$$

In order for (5) to be true, both the real and the imaginary portions must separately be equal to zero. If (5) (which comes naturally from (3) with the given substitutions) is separated into its real and imaginary parts, the resulting two equations must be simultaneous and between them express the frequency and relative values which r_p and r_g must assume in order for oscillations to exist. The particular aim in the present case is to find whether values of X_4 or X_5 exist which will enable the frequency to be expressed in terms of the constants of the circuit external to the vacuum tube so that if r_p, r_g, and μ should vary, the frequency, being dependent upon the external circuit only, will remain constant.

From (5), then, the real and imaginary parts give the following two equations:

$$\begin{aligned} &-X_0[r_p(X_2+X_4)+r_g(X_1+X_5)]-\mu r_g(X_1+X_m)(X_2+X_m) \\ &\quad = -X_0X_m\mu r_g-(X_1+X_m)^2r_g-(X_2+X_m)^2r_p. \qquad (6) \\ &X_0[r_pr_g-(X_1+X_5)(X_2+X_4)]-2X_m(X_1+X_m)(X_2+X_m) \\ &\quad = -X_0X_m^2-(X_1+X_m)^2(X_2+X_4)-(X_2+X_m)^2(X_1+X_5). \qquad (7) \end{aligned}$$

There are certain mathematical rules for finding whether the desired constancy of frequency may be obtained from the conditions given by (6) and (7). Without, however, going into detail in regard to these, it is easy to see from (7) that if X_4 and X_5 have such values as to satisfy the equation:

$$\begin{aligned} 2X_m(X_1+X_m)(X_2+X_m) &= (X_1+X_m)^2(X_2+X_4) \\ &\quad +(X_2+X_m)^2(X_1+X_5) \qquad (8) \end{aligned}$$

(which is obtained by including all terms of (7) which do not contain X_0), then the frequency of oscillation is exactly that which will cause X_0 to become zero, and will remain so, no matter what values may be taken by r_p, r_g, and μ. In other words, the oscillation frequency is equal to the series resonant frequency of the tuned circuit.

It follows, then, that if the battery voltages were to vary, the frequency, being determined by the circuit elements external to the vacuum tube only, would remain constant. In regard to a changing load resistance, it is evident that if this were connected in parallel either with r_p or r_g, then the combination of the two resistances could be considered as a single resistance. It therefore follows that the same adjustment which causes the frequency to be independent of battery voltage is also the correct one to render the frequency independent of a variable load impedance when this impedance is resistive, only, and is connected in parallel either with the plate or grid resistance of the tube.

In order to complete the general demonstration, it remains to show that the values imposed on (7) by the condition of (8) do not require physically impossible values of r_p, r_g, and μ in order to satisfy (6) and thus maintain oscillation. To do this, assume that (8) is solved for either X_4 or X_5 and substitute in (6), remembering that X_0 is zero. The result is:

$$\frac{r_p}{r_g} = \mu\left(\frac{X_1 + X_m}{X_2 + X_m}\right) - \left(\frac{X_1 + X_m}{X_2 + X_m}\right)^2. \tag{9}$$

Inspection of this expression shows that the conditions required are physically possible, and it follows that the amplitude of oscillation increases or decreases until the effective values of r_p and of r_g, which are measures of the dissipation of energy on the plate and on the grid sides, take up the values specified by the conditions of (9). Thus, for instance, if X_1 and X_2 were approximately equal, then r_p would have to be $(\mu - 1)$ times as large as r_g before the oscillation amplitude settled down to a steady state value. To many who are accustomed to neglect the losses occurring on the grid side of a vacuum tube when dealing with oscillator problems, this low value of r_g will appear as somewhat unusual. In this connection, it may be pointed out that the low value of r_g is not in any way a special requirement imposed by the stabilizing reactances, X_4 and X_5, but is inherent in vacuum tube oscillators in general, unless particular conditions are arranged to render it otherwise. For instance, it is a well-known experimental fact that resistances of the order of 4000 ohms may be placed across the grid-filament terminals of an oscillator employing any of the more common types of three-element receiving amplifier tubes without stopping the oscillations, when a good low-loss tuned circuit is employed. In view of the fact that the amplitude of the oscillations is commonly limited by r_g, this is evidence that stable oscillations may be secured with values of r_g which are of the order of 2000 or 3000 ohms.

The demonstration may be made more rigid by the use of (6) and for the special case where $X_1 = X_2$ and $X_4 = X_5 = X_M = 0$, in which the stabilizing reactances have been omitted. For such a simplified circuit, it is found by elimination of X_0 between (6) and (7) that:

$$\frac{r_p}{r_g} = \mu\left[\frac{r_p r_g - X_1^2}{r_p r_g + X_1^2}\right] - 1.$$

Now, X_1 is of the order of 500 or 600 ohms at the most, while both r_p and r_g are at least enough larger than this in the case of the more commonly used vacuum tubes so that the expression for r_p/r_g is roughly equal to $(\mu - 1)$. Thus, in the simplest kind of vacuum tube circuit, it is seen that r_g is apt to be appreciably smaller than r_p, and by no means negligible in its effect.

To return to (8), which expresses the relation between X_4, X_5 and other circuit reactances which are necessary to cause the frequency to be independent of battery voltages, we note that, although (8) is still in generalized form, and is yet to be applied to the particular case shown in Fig. 2, the very significant fact that the oscillation frequency for this type of stability must be the series resonant frequency of the tuned circuit is a direct consequence of the requirements of the equation.

For application to the Hartley type of oscillator, the various terms of (8) have the following significance:

$$X_1 = \omega L_1,$$
$$X_2 = \omega L_2,$$
$$X_m = \omega M,$$

where $\omega = 2\pi \times$ frequency and X_4 or X_5 are to be determined. In the case of Fig. 2, where stabilization is accomplished on the plate side, we put X_4 equal to zero. Then solving (8) for X_5, we find:

$$X_5 = 2\omega M\left(\frac{L_1 + M}{L_2 + M}\right) - \omega L_2\left(\frac{L_1 + M}{L_2 + M}\right)^2 - \omega L_1.$$

X_5 is thus required to be negative, so that a capacitive reactance is necessary for plate stabilization of a Hartley-type oscillator. Thus putting

$$X_5 = -\frac{1}{\omega C_5}$$

and remembering that since $X_0 = 0$, the angular frequency is given by

$$\omega^2 = 1/C_3(L_1 + L_2 + 2M),$$

finally we get

$$C_5 = C_3 \frac{L_1 + L_2 + 2M}{L_1 + L_2\left(\frac{L_1 + M}{L_2 + M}\right)^2 - 2M\left(\frac{L_1 + M}{L_2 + M}\right)}, \tag{10}$$

which is the value of capacity which should be inserted between the plate and the tuned circuit of a Hartley-type oscillator in order to cause the frequency to remain constant when the battery voltages are varied, and there is no reactance between the grid and tuned circuit.

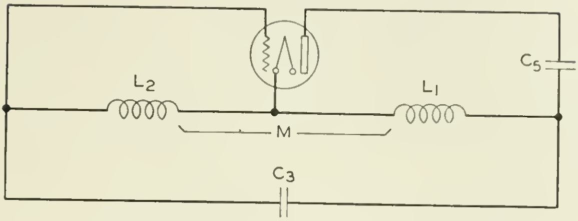


Fig. 2—Hartley oscillator, plate stabilization.

$$C_5 = C_3\left[\frac{L_0}{L_1 + L_2A^2 - 2MA}\right],$$

where

$$L_0 = L_1 + L_2 + 2M, \qquad A = \frac{L_1 + M}{L_2 + M}.$$

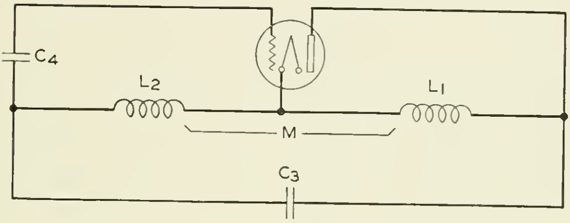


Fig. 3—Hartley oscillator, Grid stabilization.

$$C_4 = C_3A^2\left[\frac{L_0}{L_1 + L_2A^2 - 2MA}\right],$$

where

$$L_0 = L_1 + L_2 + 2M, \qquad A = \frac{L_1 + M}{L_2 + M}.$$

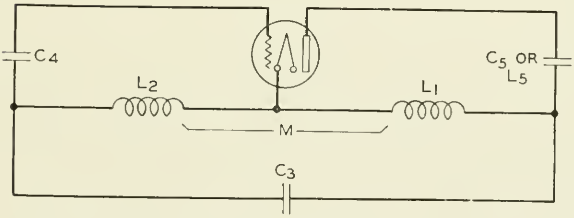


Fig. 4—Hartley oscillator, plate and grid stabilization.

$$\frac{1}{C_5} + \frac{A^2}{C_4} = \frac{1}{C_3}\left[\frac{L_1 + L_2A^2 - 2MA}{L_0}\right],$$

$$L_5 = L_0\frac{C_3}{C_4}A^2 - L_1 - L_2A^2 + 2MA,$$

where

$$L_0 = L_1 + L_2 + 2M, \qquad A = \frac{L_1 + M}{L_2 + M}$$

Of course in actual practice, it is necessary to provide a d-c path for the space current of the tube, which can be accomplished by shunting the condenser, C_5, with a high impedance choke.

It is often the case that a stopping condenser is desirable in the X_4 position, instead of a direct connection between grid and tuned circuit. This stopping condenser and the accompanying leak are advantageous inasmuch as it has been found by experience that an oscillator operating with a leak and condenser combination is inherently much more stable as regards change of frequency with change of battery voltage than an oscillator with a d-c low resistance path from grid to filament, even when a battery is employed to impose a negative bias on the grid. The explanation for this improved stability lies in the fact that the grid leak tends to keep the grid resistance, r_g, constant. It frequently happens, when the leak and condenser combination is used, that difficulty is experienced in avoiding "blocking" when a large enough condenser to have negligible reactance is employed. In such cases the required value of C_5 may be chosen in the manner indicated in connection with Fig. 4, which allows for a finite reactance between grid and tuned circuit, or else, as another alternative, the plate may be directly connected to the tuned circuit so that X_5 is zero, and stabilization may be accomplished by choosing the value of C_4 in accordance with the requirements then imposed by (8), which refer to Fig. 3 and necessitate the value of capacity shown in the figure. Another possible stabilizing arrangement is shown in Fig. 4, where either capacity or inductance may be used on the plate side, depending on the value of capacity at C_4 on the grid side. Yet another possible modification of Fig. 4 would be to use an inductance on the grid side. This would require a very small capacity on the plate side, and probably is less convenient than the arrangement indicated in the figure.

In all three of the cases considered thus far, the equations show that the value of the stabilizing capacity or inductance depends upon the values of L_1, L_2, M, and C_3 so that if the frequency of the oscillator were varied intentionally, by changing L_1, for instance, then a different value of stabilizing capacity or inductance would be required to secure independence of frequency and battery voltage at the new frequency. If, however, the circuit were so constructed that M were zero, and L_1 and L_2 were made so that they remained always equal to each other, then the value of the stabilizing element would depend upon C_3 only, and the frequency could be changed by varying L_1 and L_2 simultaneously without destroying the stabilizing adjustment.

COLPITTS OSCILLATOR

This property may be utilized to even greater advantage in connection with the Colpitts type of oscillator, which is illustrated in Figs. 5, 6, and 7 and will now be investigated with the aid of (8) in the same manner in which the relations necessary for stabilizing the Hartley oscillator were secured. Thus, for the Colpitts circuit:

$$X_1 = -\frac{1}{\omega C_1},$$

$$X_2 = -\frac{1}{\omega C_2}.$$

$$X_m = 0,$$

$$\omega^2 = \frac{1}{L_3}\left(\frac{1}{C_1} + \frac{1}{C_2}\right).$$

The values of the stabilizing elements required by (8) are given in Figs. 5, 6, and 7 which show several arrangements for the stabilizing impedance, as applied to the Colpitts-type oscillator. In particular, Fig. 7 shows a choice of either an inductance or a capacity on the grid side.

In Figs. 5 and 6 and in Fig. 7 when inductance is used on both plate and grid sides it is evident that if the condensers, C_1 and C_2, are connected together in a "gang" mounting so that when they are varied, the ratio of their capacities remains constant, then the frequency of the oscillator may be changed by changing C_1 and C_2 without disturbing the stabilizing adjustment which causes the frequency to be independent of battery voltages.

FEED-BACK OSCILLATOR

Figures 8, 9, and 10 show conventional drawings of the type of oscillator circuit known as a "feed-back" or sometimes as a "tuned input" circuit. In Fig. 8 stabilizing is accomplished on the plate side; in Fig. 9 on the grid side; and in Fig. 10 on both sides. A mathematical analysis analogous to that which was described in detail in connection with Fig. 1 gives the values of stabilizing impedances which are shown in the figures, and also indicates that the conditions for oscillation may be met when these values are employed.

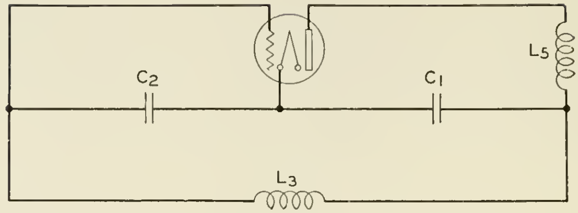


Fig. 5—Colpitts oscillator, plate stabilization.

$$L_5 = L_3 \frac{C_2}{C_1}.$$

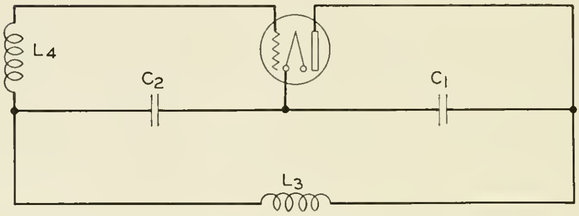


Fig. 6—Colpitts oscillator, grid stabilization.

$$L_4 = L_3 \frac{C_1}{C_2}.$$

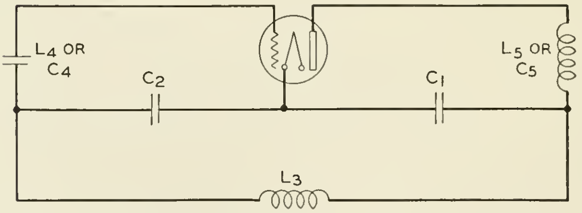


Fig. 7—Colpitts oscillator, plate and grid stabilization.

$$L_4\left(\frac{C_2}{C_1}\right) + L_5\left(\frac{C_1}{C_2}\right) = L_3,$$

$$L_5 = L_3 \frac{C_2}{C_1}\left[1 + \frac{C_2}{C_4}\left(\frac{C_2}{C_1 + C_2}\right)\right].$$

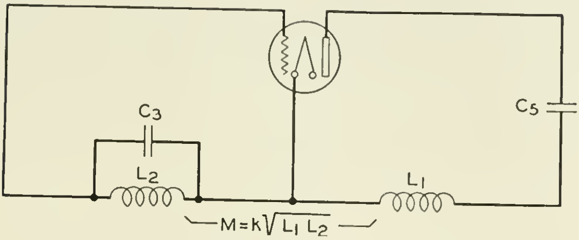


Fig. 8—Feed-back oscillator, plate stabilization.

$$C_5 = C_3 \frac{L_2}{L_1}\left(\frac{1}{1 - k^2}\right).$$

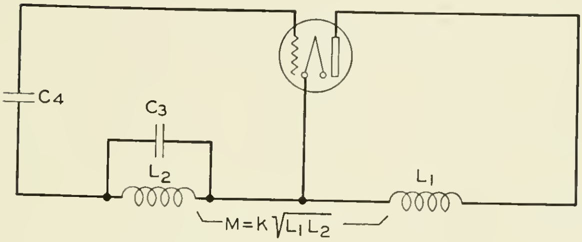


Fig. 9—Feed-back oscillator, grid stabilization.

$$C_4 = C_3\left(\frac{k^2}{1 - k^2}\right).$$

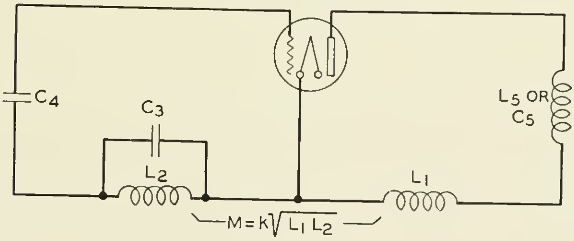


Fig. 10—Feed-back oscillator, plate and grid stabilization.

$$L_5 = L_1\left[k^2\left(1 + \frac{C_3}{C_4}\right) - 1\right],$$

$$C_5 = C_3 \frac{L_2}{L_1}\left[\frac{1}{1 - k^2\left(1 + \frac{C_3}{C_4}\right)}\right].$$

Reversed Feed-Back Oscillator

Figures 11, 12, and 13 show conventional drawings of the type of oscillator circuit known as a "reversed feed-back" or sometimes as a "tuned output" type of oscillator, with the application of stabilizing

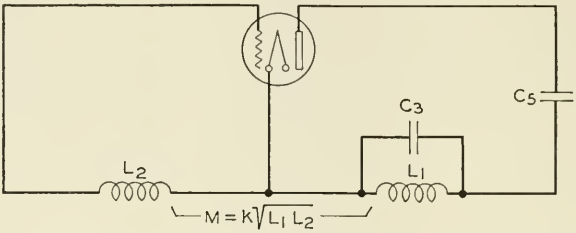


Fig. 11—Reversed feed-back oscillator, plate stabilization.

$$C_5 = C_3\left(\frac{k^2}{1-k^2}\right).$$

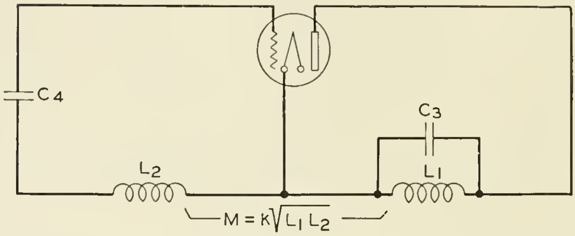


Fig. 12—Reversed feed-back oscillator, grid stabilization.

$$C_4 = C_3\frac{L_1}{L_2}\left(\frac{1}{1-k^2}\right).$$

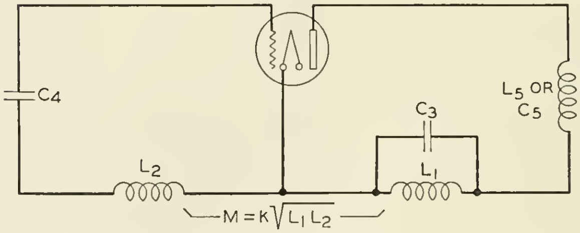


Fig. 13—Reversed feed-back oscillator, plate and grid stabilization.

$$L_5 = L_1\left[1 + \frac{1}{k^2}\left(\frac{L_1C_3}{L_2C_4} - 1\right)\right],$$

$$C_5 = \frac{C_3}{\frac{1}{k^2}\left(1 - \frac{L_1C_3}{L_2C_4}\right) - 1}.$$

impedances to cause the frequency to be independent of changes in battery voltages. In Fig. 11 the stabilizing impedance is placed between the plate and the tuned circuit; in Fig. 12 between the grid and coupling coil; and in Fig. 13 stabilization is accomplished by impedances placed in both positions. Again, the mathematical analysis gives the values of the stabilizing impedances as shown on the figures and indicates that oscillation is possible when these values are used.

OTHER TYPES OF OSCILLATOR CIRCUITS

As an instance of the stabilizing of another general class of oscillator circuits which has wide application in a number of special cases, attention is drawn to the tuned-plate, tuned-grid circuit of Fig. 14. The input and output circuits are shown in the drawing as consisting of condenser and inductance combinations connected in parallel. At any specified frequency, however, the parallel combination may be replaced by a series circuit, or, in fact, by any form of network which has the same impedance, and none of the currents or voltages in the remainder of the circuit will be altered. In particular, the inductance and capacity shown on the input side in Fig. 14 may be replaced by a piezo-electric crystal, as shown in Fig. 14-a, having the same impedance at

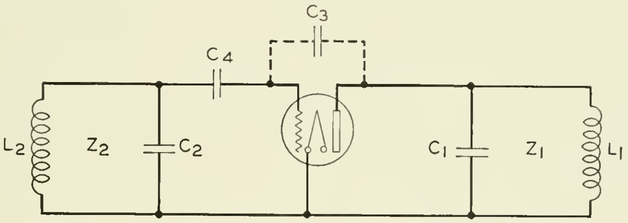


Fig. 14—Tuned-plate, tuned-grid oscillator with no magnetic coupling.

$$C_1 = \frac{L_2}{L_1}\left[C_2 + \frac{(1+\mu)C_3C_4}{C_4 + (1+\mu)C_3}\right] - C_3.$$

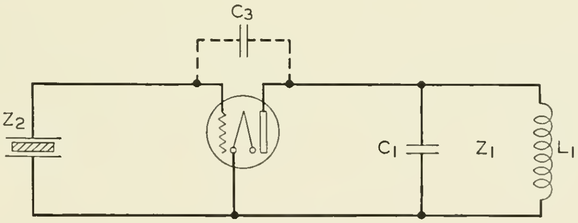


Fig. 14-a.

the operating frequency without affecting the currents and voltages in the remaining parts of the circuit.

It is well known that the frequency of such a piezo-electric oscillator is less affected by changes in battery voltage than is the frequency of the ordinary, nonstabilized electric oscillator. However, the battery voltage does influence the frequency of the piezo-electric oscillator to an extent which is undesirable for certain accurate types of work. It therefore becomes useful to apply stabilization to the piezo-electric oscillator. It will be shown that the stabilization may be accomplished by adjusting the size of the output tuning condenser to such a value that the impedance of the output circuit bears a certain critical relation to the impedance of the crystal, while at the same time, the circuit as a whole fulfills the conditions necessary for the existence of oscillations.

The same kind of stabilization is, of course, applicable to an electric oscillator having analogous relations between the input and output impedances. Thus, it is often possible to stabilize the Hartley oscillator by moving the connection between the filament and coil to different positions on the coil, until that one which gives the proper ratio of input to output impedances has been found. In the case of the Hartley and Colpitts oscillators, however, it is more often preferable to stabilize by the special circuit arrangements illustrated in Figs. 1 to 7, while, on the other hand, the tuned-grid, tuned-plate type of circuit lends itself readily to stabilization by adjustment of the output circuit.

Numerical expressions for the proper impedance relations may be obtained by noting that the circuit of Fig. 14 may be generalized into the circuit of Fig. 1 by regarding Z_4 and Z_5 as zero, while Z_2 comprises the whole input network which may consist of various arrangements of coils, condensers, grid leaks, and the like, and, in a similar fashion, Z_1 comprises the whole output network. The mathematical analysis given in connection with Fig. 1 may therefore be adapted to fit Fig. 14 immediately, and in place of (6) and (7) we have the two expressions:

$$X_0(r_g X_1 + r_p X_2) + \mu r_g X_1 X_2 = r_p X_2^2 + r_g X_1^2, \qquad (11)$$

$$X_0(r_p r_g - X_1 X_2) = - X_1 X_2 (X_1 + X_2). \qquad (12)$$

The requirements of (11) are:

$$r_p = r_g \frac{X_1}{X_2}\left[\frac{\mu X_2 + X_0 - X_1}{X_2 - X_0}\right], \qquad (13)$$

which may be used to eliminate r_p in (12) and gives

$$X_0 r_g^2(\mu X_2 + X_0 - X_1) = X_2^2(X_0 - X_1 - X_2)(X_2 - X_0). \quad (14)$$

In order for the frequency to be independent of r_g, it is necessary for one of the factors on the left-hand side of the equation to be zero. This, however, necessitates that one of the factors on the right-hand side of (14) also should be zero. Investigation shows that the only pair of factors of (14) that may both be zero, and still be consistent with (13) is the following:

$$\mu X_2 + X_0 - X_1 = 0, \quad (15)$$

$$X_2 - X_0 = 0. \quad (16)$$

Elimination of X_0 between these two expressions results in the following relation:

$$(1 + \mu)X_2 = X_1. \quad (17)$$

The frequency is then given by the expression:

$$X_1 + X_3 = 0. \quad (18)$$

Equation (17) expresses the relation which is required between the reactances of the input and the output network in order to provide for a constant frequency with varying battery voltages.

In the application of this stabilization to a piezo-electric oscillator such as is shown in Fig. 14-a it sometimes happens that stability improves with decrease in the value of the output tuning capacity but oscillations cease before complete stabilization is secured. The explanation for this and its remedy may be obtained from (17) and (18) by supposing that the reactance, X_2, of the crystal may be represented by an antiresonant circuit, C_2 and L_2, in series with a capacity, C_4, while the output reactance, X_1, consists of the antiresonant circuit, C_1 and L_1. Thus, the value of C_1 which satisfies (17) and (18) is

$$C_1 = \frac{L_2}{L_1}\left[C_2 + \frac{(1 + \mu)C_3}{C_4 + (1 + \mu)C_3}C_4\right] - C_3. \quad (19)$$

For discussion, the form which (19) takes in the absence of the stopping condenser, C_4, is:

$$C_1 = \frac{L_2}{L_1}[C_2 + (1 + \mu)C_3] - C_3.$$

This shows that a fairly large value of C_1 may be required when C_4 is absent, which places the tuning of the plate antiresonant circuit in a rather critical portion of its reactance characteristic. In order to avoid this, the introduction of a fairly small condenser at C_4 is advantageous. Thus, if C_4 were made somewhat smaller than C_3, then the value of C_1 required by (19) is roughly:

$$C_1 = \frac{L_2}{L_1}[C_2 + C_4] - C_3$$

which gives an appreciably smaller value of C_1 and results in stabilization with a much less critical adjustment than is the case when the stopping condenser is absent.

In all of the above analyses, the requirement of a capacity or an inductance is indicated by the fact that the signs come out right in the final equations. If the wrong type of reactive element were used, it would result, for example, that a negative inductance apparently would be required, which of course would indicate the requirements of a capacitance.

ANOTHER TYPE OF STABILIZATION

A third general type of stabilization may be illustrated by considering a hypothetical oscillator having its plate circuit coupled back to its grid circuit by means of a transformer coil with a coefficient of coupling equal to unity. Methods of obtaining the equivalent effect of such a coil under practical operating conditions will be described later, so that for the present it will be assumed that a unity coupled coil is on hand. The equivalent circuit diagram of the oscillator is shown in Fig. 15, where the primary and secondary windings of the coil are indicated at X_1 and X_2 respectively.

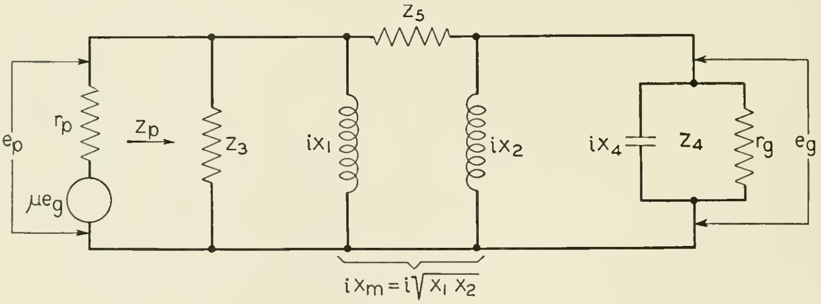


Fig. 15—Equivalent circuit of oscillator with unity coupling.

From the properties of unity coupled coils it follows that, no matter what impedances are hung across the coil, or connected between windings, the ratio of the voltage across the secondary to the voltage across the primary depends upon the coil reactances only, and not at all upon the attached impedances. In the circuit of Fig. 15 this ratio is given by the expression:

$$\frac{e_g}{e_p} = -\sqrt{\frac{X_2}{X_1}}. \tag{20}$$

In general, the voltage e_p may be expressed in terms of the impedance looking out of the plate-filament terminals of the tube. Thus

$$e_p = -\frac{\mu e_g Z_p}{r_p + Z_p} \tag{21}$$

where Z_p is the aforementioned impedance.

From (20) and (21) there results:

$$\frac{1}{Z_p} = \frac{1}{r_p}\left[\mu\sqrt{\frac{X_2}{X_1}} - 1\right]. \tag{22}$$

This equation completely expresses the operation of the oscillator in so far as impedance relations for the fundamental current component are concerned. From Fig. 15 ordinary circuit analysis shows that Z_p may be written

$$\frac{1}{Z_p} = \frac{1}{Z_3} + \frac{1}{Z_4}\left(\frac{X_2}{X_1}\right) + \frac{1}{iX_1} + \frac{1}{Z_5}\left(1 + \sqrt{\frac{X_2}{X_1}}\right)^2$$

so that (22) becomes:

$$\frac{1}{Z_3} + \frac{1}{Z_4}\left(\frac{X_2}{X_1}\right) + \frac{1}{iX_1} + \frac{1}{Z_5}\left(1 + \sqrt{\frac{X_2}{X_1}}\right)^2 = \frac{1}{r_p}\left[\mu\sqrt{\frac{X_2}{X_1}} - 1\right]. \tag{23}$$

The next step is to separate this into its real and imaginary components. We stipulate, as in the previous analyses, that the losses in the external circuit elements are small compared with those occasioned by the grid resistance of the tube. With this understanding, Z_4 may be separated into two parts, the one iX_4, in parallel with the other which constitutes the grid resistance r_g. Both Z_3 and Z_5 are taken as pure reactances. Thus the last expression yields the two equations:

$$\frac{1}{X_3} + \frac{1}{X_4}\left(\frac{X_2}{X_1}\right) + \frac{1}{X_1} + \frac{1}{X_5}\left(1 + \sqrt{\frac{X_2}{X_1}}\right)^2 = 0, \tag{24}$$

$$\frac{1}{r_g}\left(\frac{X_2}{X_1}\right) = \frac{1}{r_p}\left[\mu\sqrt{\frac{X_2}{X_1}} - 1\right]. \tag{25}$$

Equation (24) contains neither r_p, r_g, nor μ, so that the important conclusion can be drawn that the frequency of an oscillator with unity coupling between the plate and grid circuits depends only upon the inductances and capacities in the circuit, and not at all upon the tube parameters, r_p, r_g, and μ; provided, however, that the losses in the external circuit are small, and the harmonic voltages across the tube are small enough to allow r_p and r_g to be considered as pure resistances. In this connection, the interelectrode capacities may be grouped with the external circuit elements forming X_3, X_4, and X_5, so that no high-frequency difficulty is to be anticipated from them.

Equation (25) contains the relation between r_p, r_g, and μ necessary to insure the presence of oscillation. In practice, the amplitude of the oscillations builds up until this relation is satisfied.

The foregoing theory of the action of a unity coupled oscillator has led to an extremely useful and desirable result, namely, the independence of frequency and operating voltages. The point now remaining to be shown is how to get the unity coupling.

In attacking this question, the first thing to notice is that our theory does not require that the unity coupling condition,

$$M = \sqrt{L_1 L_2}$$

should be obtained. What actually is required is the much less rigid condition:

$$X_m = \sqrt{X_1 X_2} \tag{26}$$

where X_1 and X_2 are not limited to inductance alone.

Thus, imagine one of the impedances, say X_2, to consist of a coil L_2, in series with a condenser, C_2. We have then, by (26):

$$\omega^2 M^2 = \omega L_1 \left(\omega L_2 - \frac{1}{\omega C_2} \right) \tag{27}$$

or, writing

$$M = k\sqrt{L_1 L_2}$$

where k may now be less than one, we have from (27)

$$C_2 = \frac{1}{\omega^2 L_2 (1 - k^2)} \tag{28}$$

which gives the value of C_2 necessary to provide "unity coupling" at the operating frequency.

The value of X_2/X_1 is thus

$$\frac{X_2}{X_1} = \frac{\omega L_2 - \dfrac{1}{\omega C_2}}{\omega L_1} = \frac{\omega L_2 - \omega L_2(1 - k^2)}{\omega L_1} = k^2 \frac{L_2}{L_1}. \tag{29}$$

In practice, X_3, X_4, and X_5 would usually be capacities, to correspond to the circuit of Fig. 17. With this arrangement, and the relation given by (29) we have the frequency from (24):

$$\omega^2 = \frac{1}{L_1\left[C_3 + k^2C_4\frac{L_2}{L_1} + C_5\left(1 + k\sqrt{\frac{L_2}{L_1}}\right)^2\right]}. \tag{30}$$

The value of C_2 is thus written from (28) and (29) as follows:

$$C_2 = \left(\frac{L_1}{L_2}\right)\left[\frac{C_3 + k^2C_4\frac{L_2}{L_1} + C_5\left(1 + k\sqrt{\frac{L_2}{L_1}}\right)^2}{1 - k^2}\right]. \tag{31}$$

This is the general value of C_2 needed to stabilize the oscillator, and applies to any grid-stabilized oscillator where the unity coupling concept can be employed. In the case where C_5 and C_4 are small enough to be neglected we have the equivalent circuit of the reversed feed-back oscillator of Fig. 12, and for the value of the stabilizing capacity:

$$C_2 = \frac{L_1}{L_2}\,\frac{C_3}{1 - k^2}. \tag{32}$$

When the notation of Fig. 17 is reconciled with that of Fig. 12, this is in agreement with the conclusion reached for the reversed feed-back oscillator by the former method of analysis.

The present analysis has the twofold advantage of allowing the interelectrode capacities to be included, which results in (31) instead of (32); and of giving a more readily interpreted picture of the relation required for stability, namely the "unity coupling" condition of equation (26). Equation (31) is moreover applicable to the tuned-plate, tuned-grid type of oscillator, when there is magnetic coupling between the input and output circuits. Thus, in the particular instance when L_1 and L_2 are equal, as also are C_3 and C_4, we have from (31):

$$C_2 = C_3\frac{(1 + k^2)}{(1 - k^2)} + C_5\frac{(1 + k)}{(1 - k)}. \tag{33}$$

Hence, if tuning is done by "ganging" C_3 and C_4 together and varying them simultaneously, the stability may be maintained for all frequencies by making C_2 to consist of two parts: the one a fixed capacity equal to

$$C_5\frac{(1 + k)}{(1 - k)}$$

and the other a variable capacity "ganged" together with C_3 and equal to

$$C_3 \frac{(1 + k^2)}{(1 - k^2)} .$$

In order to insure fulfillment of the requirements of this theory, it was mentioned above that the oscillator should be relatively free from harmonics, so that r_p and r_g may be taken as pure resistances. That this requirement can be successfully met may be demonstrated by reference to Fig. 15 and consideration of the means which would be employed if the circuit represented an amplifier with e_g impressed on the grid of a following tube instead of on the grid of the driving tube itself. In such an arrangement, it is well known that the distortion is least when the impedance looking out of the plate of the driving tube is made materially larger than the internal plate resistance of the tube itself; and, second, when the impedance looking back out of the grid of the driven tube is made materially smaller than the internal grid resistance of the tube itself. The conditions which determine how nearly these two requirements may be met in the oscillator tube are governed by (22). Thus

$$\frac{r_p}{Z_p} = \mu \sqrt{\frac{X_2}{X_1}} - 1. \tag{34}$$

This should be small in order that the first requirement mentioned above should be fulfilled. Hence $\mu k \sqrt{L_2/L_1}$ should exceed unity by as little as is consistent with reliable oscillation. When we come to consider the second of the requirements for decreasing harmonics we obtain an expression analogous to (34), namely:

$$\frac{Z_g}{r_g} = \mu \sqrt{\frac{X_2}{X_1}} - 1 \tag{35}$$

so that the requirement that the impedance looking back out of the grid should be less than the grid resistance is satisfied by the same condition as that required by (34); namely, that $\mu k \sqrt{L_2/L_1}$ should exceed unity by as little as is consistent with reliable oscillation.

From Fig. 17, by regarding C_3 and C_5 as zero, the feed-back type of oscillator is obtained. Again, when C_3 and C_4 are zero, the Hartley oscillator results. The stabilization both of the feed-back oscillator and the Hartley type by the method shown on the figure was not described in the analysis of Figs. 3 and 9 inasmuch as the present method places the stabilizing element directly in the tuned circuit, whereas the former method placed it between the tuned circuit and the vacuum tube.

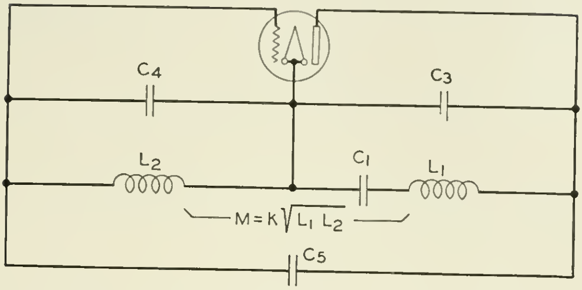


Fig. 16—General circuit of oscillator with unity coupling, plate stabilization.

$$C_1 = \frac{L_2}{L_1}\left[\frac{C_3k^2\frac{L_1}{L_2} + C_4 + C_5\left(1 + k\sqrt{\frac{L_1}{L_2}}\right)^2}{1 - k^2}\right].$$

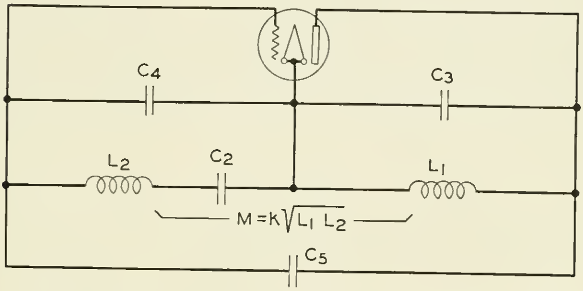


Fig. 17—General circuit of oscillator with unity coupling, grid stabilization.

$$C_2 = \left(\frac{L_1}{L_2}\right)\left[\frac{C_3 + k^2C_4\left(\frac{L_2}{L_1}\right) + C_5\left(1 + k\sqrt{\frac{L_2}{L_1}}\right)^2}{1 - k^2}\right].$$

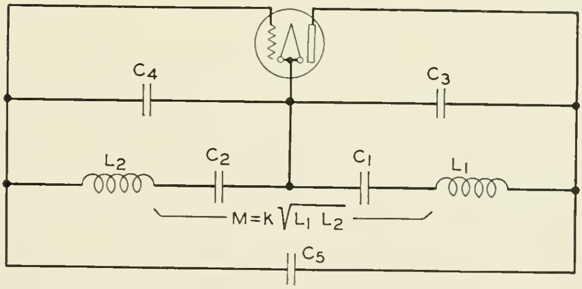


Fig. 18—General circuit of oscillator with unity coupling, plate and grid stabilization.

$$\omega^2k^2L_1L_2 = \left(\omega L_1 - \frac{1}{\omega C_1}\right)\left(\omega L_2 - \frac{1}{\omega C_2}\right).$$

Besides the general circuit of Fig. 17, where stabilization is accomplished by imposing a critical value on C_2, there is the alternative shown in Fig. 16, where the stabilizing element is C_1, a condenser located in series with L_1, and of such a size as to cancel the leakage reactance between L_1 and L_2. The formula giving this size in terms of the other circuit constants is shown in the figure, and the condition that the harmonic content should be small for circuits of the general type of Fig. 16 is that $(\mu/k)\sqrt{L_2/L_1}$ should exceed unity by as little as is consistent with reliable oscillation.

From Fig. 16 the circuits of various types of oscillators may be derived in the manner which was described in connection with Fig. 17. Of these circuits, the reversed feed-back and the Hartley were not described in connection with Figs. 11 and 12.

A combination of the features of Figs. 16 and 17 may be employed in the manner shown in Fig. 18 where condensers C_1 and C_2 are placed in series both with L_1 and L_2, respectively. In this case, the formula for the required size of C_1 and C_2 becomes quite cumbersome when expressed in terms of the other circuit elements, only. Since, however, the frequency is usually known approximately, we may use the relation

$$X_m^2 = X_1 X_2$$

for finding C_1 and C_2 in terms of ω and get:

$$\omega^2 k^2 L_1 L_2 = \left(\omega L_1 - \frac{1}{\omega C_1}\right)\left(\omega L_2 - \frac{1}{\omega C_2}\right). \tag{36}$$

As in the case of Figs. 16 and 17, so also may the circuit of Fig. 18 be modified to correspond to the reversed feed-back, the feed-back and the Hartley types of oscillators where stabilization is accomplished both on the plate and on the grid sides.

EXPERIMENT

Of course, Figs. 1 to 18 are intended to represent only the fundamentals of the corresponding circuits. For practical operation these circuits would have to include the usual stopping condensers, leak resistances, sources, choke coils, and accessories. These circuit elements should be so valued and introduced into the oscillator circuit as a whole as not to interfere with the relations required by the analyses, in order to maintain the stabilizing effects of the stabilizing impedances. As to the choke coils, this means merely that they must be what the name implies, that is, a substantially infinite impedance. In the case of a Hartley-type oscillator, where the reactance is chosen to be located in the grid leads instead of in the plate leads, a condenser must be used. This may replace the conventional stopping condenser. Where the reactance is in the plate lead for a similar type of oscillator, the stopping condenser in the grid lead should be large so as to have negligible impedance. Similar expedients are suggested for the impedances of the other types of oscillator circuits.

As typical of the general method whereby any of the simplified circuits of Figs. 1 to 18 may be elaborated into a conventional circuit of this kind, including the various adjunctive circuits, Fig. 19 should be referred to. This figure illustrates a complete wiring diagram of the oscillator of Fig. 7 and shows an example of stabilization by means of the inductance L_5 in the plate circuit and the inductance L_4 in the grid circuit. In addition to satisfying the relation shown in Fig. 7, it may be noticed that the value of L_5 is such as to tune with C_1 to the oscillation frequency, and, similarly, the value of L_4 is such as to tune with C_2 to the oscillation frequency. Under such conditions a resistance of appreciable value may be introduced into the circuit of L_3 without affecting the frequency of the stabilization. The reason for this may be explained briefly as follows:

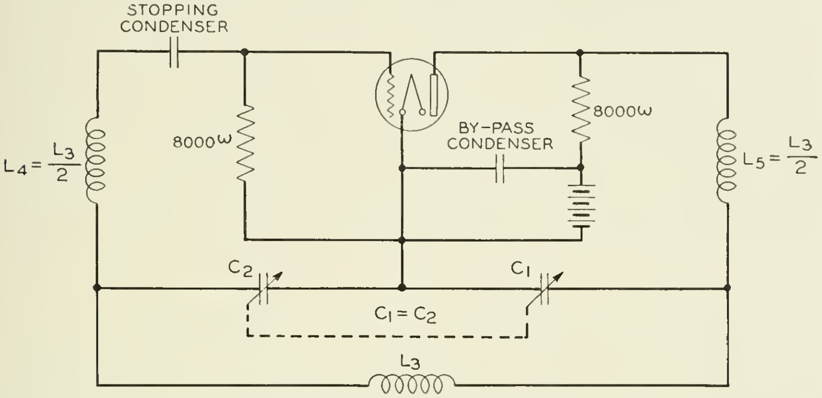


Fig. 19—Practical arrangement of oscillator with three series tuned coupled circuits.

Consider a single series circuit formed of one of the three meshes of Fig. 19, for instance that composed of the elements, r_g, in parallel with the 8000-ohm leak, L_4, and C_2. This circuit is in series resonance at the frequency at which the circuit as a whole oscillates. Therefore it tends to introduce resistance impedance only into whatever circuits it is reactively coupled with. Thus, the effect of this circuit upon the adjacent circuit, L_3, C_1, C_2, with which it is coupled is to introduce resistance only. Similarly, if this last circuit operates at series resonance, only resistance is introduced into the plate circuit, L_5, C_1, and r_p, in parallel with the d-c feed of 8000 ohms, with which it is coupled. Hence, if the plate circuit likewise operates at series resonance, a change in resistance of any part of the circuit will change only the resistance into which the tube works and therefore will leave the frequency unaltered.

In a more general sense, any of the oscillator forms discussed may be stabilized even when the resistance in the external circuit is not inappreciable, the effect of the external resistance manifesting itself in two ways: first, a value of stabilizing reactance slightly different from that given in the above formulas may be required, and second, the frequency, instead of being absolutely independent of battery voltage variations, goes through a maximum or a minimum as the battery voltages change, the voltage at which this maximum or minimum occurs depending upon the exact value of the stabilizing reactance. An exact mathematical analysis of this more general case yields formulas for the stabilizing reactances which involve r_g or r_p and hence are not as useful even in a case where the resistance in the external circuit is of importance as are the formulas presented above. The latter may be used as first approximations in any event.

In practice it has been found that when ordinary precautions are taken to insure a low-loss external circuit, the relations given above hold very accurately and any variations in frequency then existing as a result of varying battery voltages may be traced to either one of two causes, both of which may be guarded against: first, the interelectrode capacities of the tube may be sufficient to enter into the impedance relations. In this event, a change in the form of circuit, such as the use of the tuned plate-tuned grid arrangement of Fig. 17, where the interelectrode capacities form a part of the external circuit, will eliminate the difficulty. Second, the harmonic currents caused by the nonlinear characteristics of the vacuum tube may introduce the effect of a reactive impedance back into the fundamental which may vary with battery voltage and so change the frequency. The remedy for this is to provide a low reactance path for the harmonics so that they have no opportunity to build up a reactive voltage across the tube, and also to use grid leaks and other such well-known devices for reducing the harmonic currents generated by the tube.

For the purpose of providing information as to the order of stability which may be expected from the several methods of stabilization outlined above, various quantitative experimental tests have been conducted. The general results of these tests may be summarized by saying that a close adherence to the theoretical requirements results in an oscillator whose frequency depends upon operating voltages to such a small extent that temperature effects become the predominating influence and special precautions must be taken in order to eliminate them before data showing the dependence of frequency on operating voltages can be obtained.

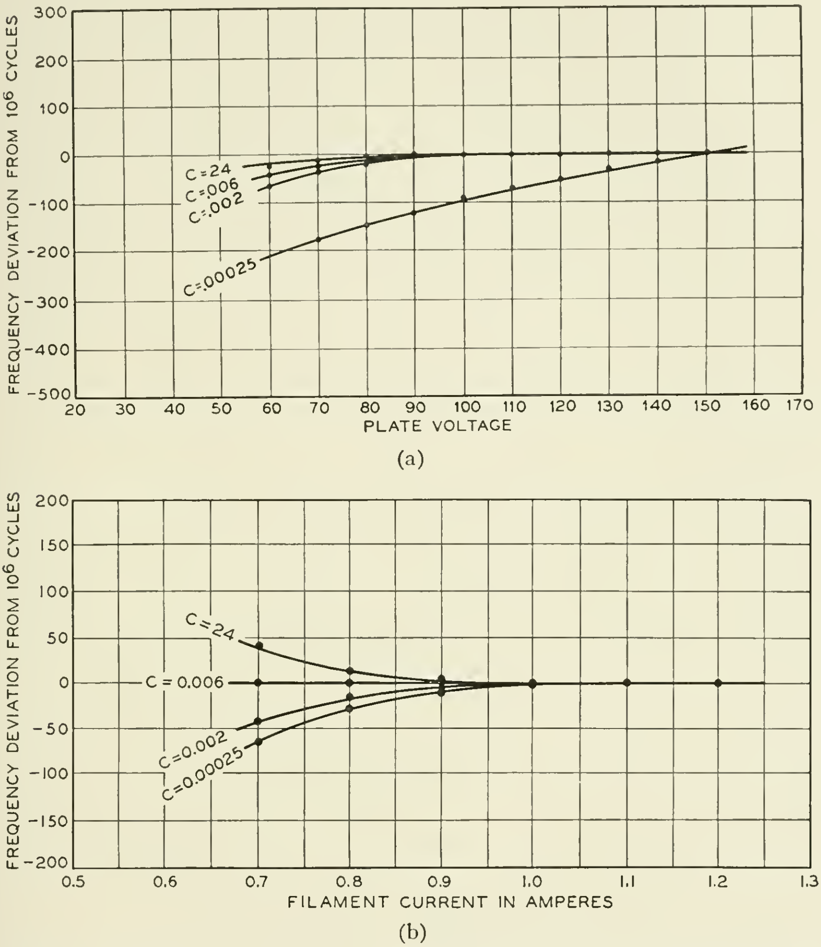
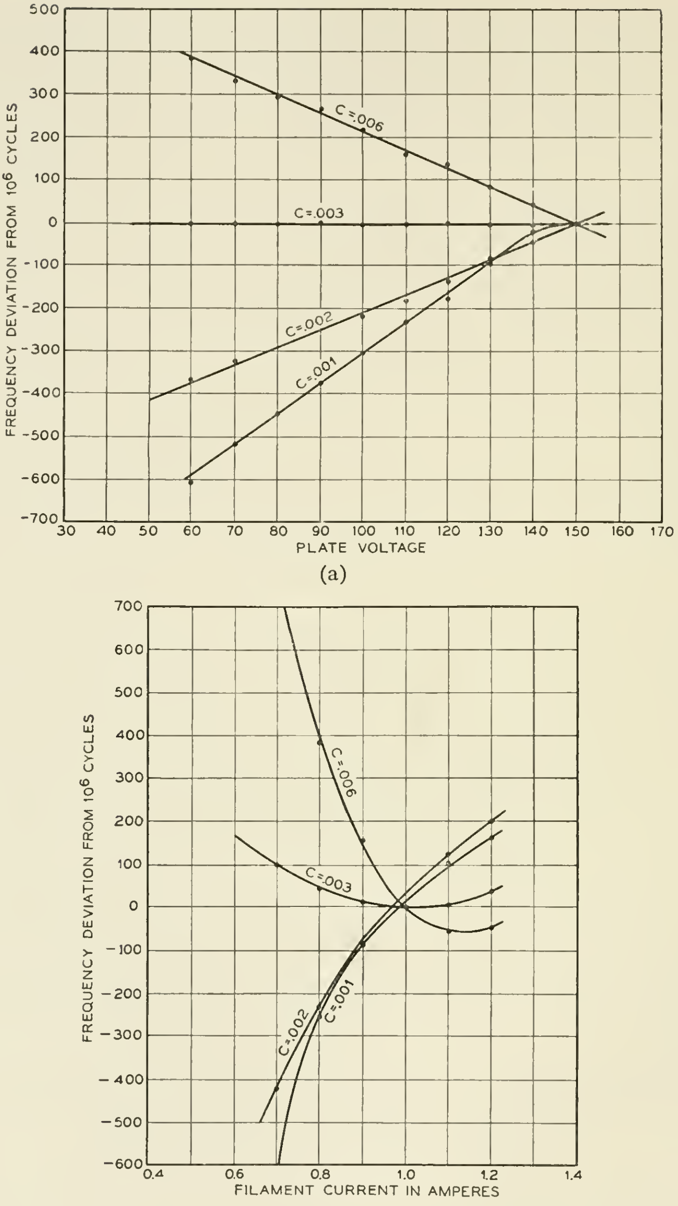


Fig. 20—Performance curves of reversed feed-back oscillator with tight coupling, grid stabilization.

(a) Variation of frequency with plate potential.
(b) Variation of frequency with filament current.

For instance, the data for the curves shown in the accompanying Figs. 20 and 21 were secured by the following procedure: A plate



(a)

(b)

Fig. 21—Performance of reversed feed-back oscillator with loose coupling, grid stabilization.

(a) Variation of frequency with plate potential.
(b) Variation of frequency with filament current.

potential of 150 volts and a filament current of one ampere were selected as reference points. The frequency under these conditions was noted. A change to a different operating condition, say 140 volts plate potential and 1 ampere filament current, was made and the frequency measured as rapidly as possible, whereupon the operating voltages were returned immediately to their reference values and the frequency rechecked. Special care was taken to keep the room temperature very constant, but, even so, the heating of the parts of the oscillator circuit by the operating alternating and direct currents was sufficient to affect the frequency to an undesirable extent, requiring that the readings be taken with unusual rapidity in order to return the voltages to normal before the changed operating currents could appreciably affect the temperature of the coils, tube elements, and other parts of the circuit.

The final results, however, are consistent enough to be representative of what can be accomplished, and the two sets of curves shown in the figures bring out a result which was found to hold throughout the investigation; namely that the higher the coefficient of coupling in the coil used to secure feed-back, the less critical was the value of the stabilizing impedance. Thus, in Fig. 20 the coupling was as tight as it was possible to produce by winding the primary and secondary simultaneously upon a tube to form a single layer solenoid, while the coil used for Fig. 21 was made by removing about half the turns from the secondary of the same coil, thus providing for a step-down in voltage as well as a decrease in coupling. A possible explanation for the less critical adjustment required with tight coupling lies in the fact that the tightly coupled coil satisfies the condition for "unity coupling" as given by (26) over a range of frequencies whereas the loosely coupled coil satisfies the condition at only the frequency critically determined by the stabilizing capacity.

This would appear to indicate that the type of stabilization described in the theoretical part of the paper under the unity coupling concept offers certain practical advantages over those types where the stabilizing element is placed between the tuned circuit and the tube, as in the Colpitts oscillator of Figs. 5, 6, and 7, for example, where no magnetic coupling whatever is employed. Experiments with the Colpitts circuit have shown, however, that when the capacities in the tuned circuit are made large relative to the inductance, a very satisfactory degree of stabilization may be secured at the lower frequencies where interelectrode capacities may be neglected. A close inspection of the theory of stabilizing the frequency of the Colpitts oscillator shows an argument analogous to that of the coupled coils, namely that the smaller we make the inductance in the tuned circuit, and the larger we make the capacity, the frequency being kept constant, the greater will be the range of frequencies over which the condition, required by (8), that the series reactance of the tuned circuit be zero, is satisfied to an approximation sufficiently good for practical purposes, and hence the less critical will be the adjustment of the stabilizing impedance.

For the data in Figs. 20 and 21 a reversed feed-back type of oscillator was employed, having an elementary circuit similar to that of Fig. 12. A grid leak was placed in parallel with the stabilizing capacity, but care was taken to see that the resistance of the leak was always so high that its value did not affect the frequency. This was done by using a variable resistance and increasing its value until the frequency no longer changed. A large grid leak has always been found advantageous in securing constancy of frequency, but the size of the leak reaches a practical limitation determined by the time constant which produces the familiar "blocking."

The plate battery potential was fed through a choke in series with the plate inductance coil, the combination of choke and B battery being thoroughly by-passed with a large condenser.

For the purpose of checking the size of the stabilizing capacity to find its agreement with theory, an indirect method was used. The theory requires that

$$C_4 = C_3 \frac{L_1}{L_2}\left(\frac{1}{1 - k^2}\right)$$

as shown in Fig. 12. The operating frequency was one megacycle, which made a direct measurement of the coupling coefficient, k, somewhat awkward, so that a method based on the "unity coupling" concept was employed. Thus from (26)

$$k^2\omega^2 L_1 L_2 = \omega L_1\left(\omega L_2 - \frac{1}{\omega C_4}\right) \tag{37}$$

may also be used for determining C_4. The primary, L_1, of the coil was connected through an impedance to a source of e.m.f. of one megacycle, and a vacuum tube voltmeter was placed across it. The condenser, C_4, was placed directly in parallel with the secondary, L_2. With this arrangement the impedance looking into the primary is

$$Z = i\omega L_1 + \frac{\omega^2 k^2 L_1 L_2}{i\left(\omega L_2 - \dfrac{1}{\omega C_4}\right)}$$

unless points very near the resonance point are considered. This last equation may be written

$$iZ\left(\omega L_2 - \frac{1}{\omega C_4}\right) = -\,\omega L_1\left(\omega L_2 - \frac{1}{\omega C_4}\right) + \omega^2 k^2 L_1 L_2. \tag{38}$$

The condenser, C_4, was varied until the reading of the vacuum tube voltmeter became zero. This means that Z was zero, and hence (38) gives the value of C_4 required by (37) which is in turn the value needed to stabilize the oscillator.

How well this checked the actual values needed for the case shown in Figs. 20 and 21 may be seen by the following: For Fig. 21 the value of C_4 measured as above, was 4000 μμf. The experimental value was 3000 μμf. For Fig. 20 the measurement gave a rather broad zero on the vacuum tube voltmeter, which was, however, estimated at 8400 μμf. For a check, a measurement was made at 7 megacycles which gave a sharper zero, and a value of 120 μμf. This must be reduced to its equivalent value of 1 megacycle by multiplying by 7^2 which gives 5880 μμf. The experimental curves of Fig. 20 show a noncritical value of 6000 μμf. which is nevertheless in good accord with the above measurements, while in Fig. 21 the agreement is somewhat more striking.

As an example of stabilization of oscillators in the altogether different frequency region from 7 to 40 kc, the following table was taken from data kindly supplied by F. J. Rassmussen:

TABLE I

Frequency kc	L_1 mh	L_2 mh	Coupling k	Stabilizing Capacity μf	
				Experimental	Theoretical
12.5	3.176	6.891	0.131	0.0021	0.0022
40.0	3.210	7.010	0.134	0.0018	0.0023
12.5	3.178	6.974	0.129	0.0021	0.0022
40.0	3.211	7.094	0.133	0.0018	0.0023
7	2.498	0.500	0.720	0.4 to ∞	2.07
15	2.499	0.498	0.723	0.2 to ∞	0.46
15	1.409	0.161	0.695	0.2 to ∞	1.33
23	1.407	0.161	0.698	0.1 to ∞	0.58
23	0.665	0.093	0.677	0.1 to ∞	0.99
31	0.665	0.093	0.681	0.1 to ∞	0.57
31	0.487	0.076	0.663	0.1 to ∞	0.64

This table again emphasizes the less critical adjustment required when the coupling is tight, as in the last seven rows.

It is hoped that the foregoing data and comments will serve as a guide to design methods for constant frequency oscillators, in so far as dependence of the frequency on operating voltages is concerned. Combinations and permutations of the various circuits dealt with will occur to the designer who requires special arrangements to fit special cases. The generalized circuit of Fig. 18 is suggested as being adaptable to meet the most widely varying conditions. This is particularly true at very high frequencies, since all the interelectrode capacities are included in the circuit of that figure.

The popular "push-pull" type of circuit may likewise be generalized to correspond to several of the fundamental circuits illustrated in the figures, and may be stabilized by the methods indicated. However, because of the nonuniformity of vacuum tubes and the added complication of the circuit, no advantage has been obtained by its use, so that the single tube circuits are to be preferred wherever special conditions do not require the push-pull type.

Appendix

The complete and rigorous mathematical relations for oscillation circuits containing vacuum tubes have seldom been discussed in connection with their practical application to useful circuits. In the case of the stabilization of oscillators against changes in battery voltages it is important to base the theory upon as strictly rigorous a mathematical foundation as possible, yet at the same time to be able to express the results in readily useful terms. It will be shown that this desirable result may be attained by a proper interpretation of the meaning of the internal impedances, r_p and r_g, of the vacuum tube.

To show this in the shortest and most obvious way, a simple series circuit will be considered. Let the circuit consist of a resistance, R, a condenser, C, and an inductance, L, all connected in series with a vacuum tube which may be taken as having a "negative resistance" characteristic. In order to increase the generality of the demonstration, a sinusoidal driving voltage, E, of angular frequency, ω, is also allowed to act on the circuit. By Kirchkoff's Law, the current in the circuit is expressed by the equation

$$E = RI + L\frac{dI}{dt} + \frac{1}{C}\int I dt + V \qquad (1)$$

where V is the drop across the vacuum tube. As a general expression for V in terms of the current the following expression may be used:

$$V = V_0 + A_1 I + A_2 I^2 + A_3 I^3 + \cdots. \qquad (2)$$

We are interested in the "steady state" solution, and accordingly a Fourier series will be the most general form which can be assumed for the current. It is convenient to write the series in the following form:

$$I = \sum_{n=-\infty}^{\infty} \frac{b_n}{2} \epsilon^{in\omega t} \tag{3}$$

or, for brevity, in the symbolic form

$$I = \sum I(n\omega) \tag{4}$$

where the summation is understood to extend from minus infinity to plus infinity. Substitution of (4) and (2) into (1) gives:

$$\begin{aligned} E = {} & R\sum I(n\omega) + L\sum in\omega I(n\omega) + \frac{1}{C}\sum \frac{I(n\omega)}{in\omega} \\ & + V_0 + A_1\sum I(n\omega) + A_2\sum\sum I(n\omega)I(m\omega) \\ & + A_3\sum\sum\sum I(n\omega)I(m\omega)I(l\omega) + \cdots. \end{aligned} \tag{5}$$

For the component of fundamental frequency, ω, we get

$$\begin{aligned} E(\omega) = {} & RI(\omega) + Li\omega I(\omega) + \frac{I(\omega)}{Ci\omega} + A_1 I(\omega) \\ & + A_2 \sum_{n+m=1} I(n\omega)I(m\omega) + A_3 \sum_{n+m+l=1} I(n\omega)I(m\omega)I(l\omega) + \cdots \end{aligned} \tag{6}$$

where the summation terms involve the products of all frequency components which beat together to give the fundamental, as indicated. In order to put the last expression in symmetrical form, it is convenient to multiply and divide each of the summation terms by $I(\omega)$ so that we may write

$$\begin{aligned} E(\omega) = I(\omega) \Bigg[& R + Li\omega + \frac{1}{Ci\omega} + A_1 + \frac{A_2}{I(\omega)} \sum I(n\omega)I(m\omega) \\ & + \frac{A_3}{I(\omega)} \sum I(n\omega)I(m\omega)I(l\omega) + \cdots \Bigg]. \end{aligned} \tag{7}$$

This expression exhibits the terms in square brackets in the form of an impedance, and shows that the vacuum tube may be treated as an ordinary linear circuit element if it is considered as having the impedance

$$Z = A_1 + \frac{A_2}{I(\omega)} \sum I(n\omega)I(m\omega) + \frac{A_3}{I(\omega)} \sum I(n\omega)I(m\omega)I(l\omega) + \cdots. \tag{8}$$

Of course, the numerical value of such an impedance cannot be found from this expression alone, but in oscillator analysis there is no necessity for its numerical evaluation. The very important fact that the nonlinear elements in a circuit network may be replaced by equivalent impedances so that the ordinary circuit analysis can be employed has been demonstrated. It is possible to tell something about the form of the tube impedance from (8). Thus, the first term, namely A_1, is a real quantity and contributes a part of the total effective resistance of the tube. All of the remaining terms are, in general, complex, depending upon the phases of the different harmonic currents. Thus, the conclusion is reached that a nonlinear resistance may be reduced to an equivalent linear impedance, but that this impedance has a reactive as well as a resistive component in the general case. There is at least one important instance where the equivalent linear impedance is resistive only. This occurs when the impedance in the circuit external to the vacuum tube contains resistance, only, to all of the harmonic currents.

With the general conception of the impedance of the vacuum tube, described above, the fundamental component of (1) becomes

$$E = \left[R + i\omega L + \frac{i}{i\omega C} + r + iX \right] I \tag{9}$$

where the tube impedance is represented by $r + iX$.

When the driving voltage, E, is zero, as in the case of oscillators, then for a finite current to exist the oscillation conditions are:

$$\left. \begin{aligned} R + r &= 0 \\ \omega L - \frac{1}{\omega C} + X &= 0 \end{aligned} \right| . \tag{10}$$

In the treatment of oscillator networks employed in the foregoing paper the quantities, r_p and r_g, are used in the sense of the resistance, r, in (9) and (10) of this appendix. The reactive component, X, of the tube impedance has been neglected in the paper, for the reason that all of the circuits discussed are of such character that the reactance of the external circuit to the harmonic currents may be made quite low, and the nonlinearity of the vacuum tube characteristics is not such as to cause excessive production of harmonics.

In the case of the dynatron type of oscillator, where the harmonic currents are especially strong, it has been found by experiment that the reactive component of the tube impedance cannot be neglected, but that it is, in fact, altogether responsible for the variation in frequency with battery voltages which is characteristic of the dynatron oscillator.

Some Physical Properties of Wiping Solders *

By D. A. McLEAN, R. L. PEEK, Jr., and E. E. SCHUMACHER

The plasticity of a number of solders at wiping temperatures has been determined by compression tests between parallel plates. The character of the flow is found to be that corresponding to a linear relation between shearing stress and a fractional power of the velocity gradient. This corresponds approximately to a relation between rate of compression (dh/dt) and sample height (h) given by the equation: $dh/dt = kh^b$, in which k and b are constants, of which b is independent of the test conditions. For viscous materials $b = 5.0$; for most solders b is greater than 5.0, and increasing values of b are associated with lower temperature gradients of plasticity. It is shown that a solder must have a low temperature gradient of plasticity in order to be properly wiped, and that determination of the value of b by means of a plasticity test can therefore be used to evaluate the working properties of a solder.

A number of factors upon which the plasticity of wiping solders and the porosity of wiped joints may depend have been investigated. In particular, it is shown that segregation is not responsible for porosity, but that the latter may be dependent upon the particle size of the solid phase at wiping temperatures. The relation of particle size to the wetting power of the liquid phase is discussed.

A WIPING solder is one used in joining sections of lead (or lead alloy) piping or sheathing, such as that used for telephone cables. As such, it must wet the sheathing readily, must be coherent and plastic enough to be worked with the hands over a considerable temperature range, and must form a strong, non-porous joint. The desirable qualities for a wiping solder have been more fully enumerated by Schumacher and Basch,[1] but the above will suffice for consideration in connection with this paper.

One phase of a general investigation of wiping solders has been a study of their plastic properties in the temperature range in which they are wiped. In this study consideration has been given not only to the working properties which a solder must have in order to be properly wiped, but also to certain other properties which may affect the character of the wiped joint, particularly those which may result in porosity of the joint.

The temperatures at which the tests described below were made lie in the wiping range, and are intermediate to those represented by the solidus and liquidus lines in the respective equilibrium diagrams of the solders investigated. At such temperatures the alloys form two phase

* Presented before the Society of Rheology, Rochester, New York, December 28, 1931. Published in *Jour. of Rheology*, January, 1931.

[1] E. E. Schumacher and E. J. Basch, *Ind. & Eng. Chem.*, **21**, 16 (1929).

systems, in which a solid phase, composed principally of one constituent, is dispersed in a liquid phase, whose composition and abundance is a function of the temperature and the gross composition of the alloy. These facts suggest that the character of a solder may be related in the wiping range to the following properties: solid-liquid ratio, solid particle size, viscosity of the liquid phase, tendency to segregation, and the interfacial tension between solid and liquid phases.

Objects of Investigation

The objects of the investigation may be summarized under the following headings:

1. Character of Flow: It was hoped that by means of plasticity determinations, data would be obtained permitting the flow of the various alloys to be formulated in simple terms, and that appropriate flow constants could be evaluated which would suffice to describe the nature and extent of deformation under specified conditions of stress. This object was largely realized through the discovery that the flow is similar in character to that of many colloidal dispersions, corresponding approximately to what would be expected if a power relation exists between stress and velocity gradient.

2. Relation of Plasticity to Workability: The primary purpose of the plasticity studies was to determine if the information thus obtained could be related to the working characteristics of the solders. If a clear-cut relation could be found, the plastometer could be employed in testing new alloys for use as solders, and in making control tests on solders as supplied to the splicers. As a matter of fact, a simple relation between the plasticity data and workability was found to exist, on which can be based a method of employing the plastometer in research studies of solders. In addition the plasticity appears to be very sensitive to composition changes, and its measurement with the plastometer should in consequence be useful as a control test. Whether this latter possibility can be realized in engineering practice has not, as yet, been ascertained.

3. Plasticity and Solid-Liquid Ratio: Early in the investigation it was suggested that the plasticity might be related to the solid-liquid ratio. If this were so, the plasticity and its temperature gradient could be predicted from the equilibrium diagram for any particular alloy, and calculation of solid-liquid ratios could be used to supplant or to corroborate plasticity determinations. While for any one solder, of course, the plasticity increases with the ratio of liquid to solid, no relation common to any group of solders was found to hold between plasticity and the solid-liquid ratio.

4. Porosity: At the start of the investigation it was thought that the porosity of the joint might be related to the plasticity of the solder. As this was not found to be the case, other factors which might be responsible for this condition were investigated. In particular, experiments were performed to determine if porosity is dependent upon segregation, or if it is related to the particle size of the solid phase at wiping temperatures. Porosity and segregation were found to be quite independent, but a relation between porosity and particle size was found in the two tests made in this connection.

Experimental work directed toward these objects has involved the following groups of experiments:

1. Plasticity studies.
2. Segregation studies.
3. Investigation of particle size.

The first of these groups was planned to cover the first three objects enumerated above, while the other two groups are of interest in connection with porosity.

ALLOYS INVESTIGATED

Six different alloys were tested as described in Table I. In this table, the compositions sought in preparing the samples are given, together with the composition found by analysis at the top and bottom of the cast (T & B), except for solders Nos. 5 and 6, of which no analyses were made.

Nos. 2 and 3 are good, workable solders giving non-porous joints, with No. 3 possibly being given preference in ease of handling while the joint is being wiped. No. 1 is also satisfactory from the standpoint of workability, but the joints formed from it are often porous, while the splicers described it as being somewhat " coarse." No. 4 is unsatisfactory in all respects. Some tests on No. 5 showed it to be good, but in recent tests the workability has been poor and a number of porous joints have been found. Solder No. 6 seemed to be fair both in workability and porosity, but the number of tests has been altogether too limited for definite conclusions to be drawn. The workability of neither of these last two solders is as good as that of the lead-tin and lead-tin-cadmium alloys. The results of wiping and porosity tests are summarized to the right of Table I.

A point which the authors wish to stress is that they are interested here only in two of a number of factors which may affect the utility of solders; namely, workability and porosity. Therefore, that a solder is good in those respects does not necessarily imply that it is a good solder in a general sense, and leaves open the question as to

TABLE I

	Per Cent Pb	Per Cent Sn	Per Cent Cd	Per Cent Bi	Per Cent Zn	Workability	Lack of Porosity	Tested at ° C.
Solder No. 1.........						Good	Porous	196, 204 and 210
Desired.........	65	35						
Actual T........		34.50						
Actual B........		34.62						
Solder No. 2.........						Good	Good	183, 185 and 194
Desired.........	60	40						
Actual T........		40.32						
Actual B........		40.69						
Solder No. 3.........						Excellent	Good	170, 184 and 192
Desired.........	67	24	9					
Actual T........		24.00	9.28					
Actual B........		24.13	9.38					
Solder No. 4.........						Poor	Porous	
Desired.........	60			40				
Actual T........				40.25				
Actual B........				40.20				
Solder No. 5.........						Fair	Porous	149,153, 157, 161
Desired.........	67		4.25	28.5	.25			
Solder No. 6.........						Fair	Good	140, 148, 156, 160
Desired.........	65	13		22				

T = Top of cast.
B = Bottom of cast.

whether or not consideration of other factors may show it to be unsatisfactory as a wiping solder.

Plasticity Studies

Experimental Procedure

The plastometer used in the investigation is a modification of the instrument used by Williams[2] in studies on rubber compounds. It has been fully described elsewhere[3] by one of the writers. A heat-insulated cylindrical steel block provided with heating elements and thermocouples contains a central cylindrical well 2½ inches (6.35 cm.) in diameter and 5 inches (12.70 cm.) deep. The block itself is about 12 inches (30.48 cm.) in diameter, being made large to prevent rapid temperature fluctuations. A flat-bottomed cylindrical plunger, also carrying heating elements and a thermocouple, fits into the well with a small clearance. The load on the plunger may be adjusted to suit the conditions of test. A sample of the material to be tested is placed under the plunger, given 90 minutes to come to temperature and the plunger released and allowed to compress the sample. The temperature is frequently checked, and can be controlled within ± 1° C. by adjusting external resistances. The sample height is measured at intervals by an Ames gauge which moves with the plunger, a complete record of sample height vs. time being thus obtained. With one exception, the tests here reported were continued for 30 minutes.

---

[2] I. Williams, *Ind. & Eng. Chem.*, 16, 262 (1924).

[3] R. L. Peek, Jr., "Parallel Plate Plastometry" (now being prepared for publication).

The sample size used was one which has been found convenient—a cylinder 1.374 inches (3.490 cm.) in diameter and .300 inch (.762 cm.) high. Samples were obtained from a single cast of each solder, tested for blow-holes and inclusions by density measurements. Samples were milled to dimensions with a tolerance of ± .001 inch (.0025 cm.). The weight of the plunger and load was 30 lbs. (13.63 kg.).

Character of the Flow

The data directly obtained in a typical run, readings of sample height vs. time from start of run, are plotted in Fig. 1, together with tangents to the curve fitting the points plotted. By computing the slope of such tangents there are obtained values of the rate of compression, dh/dt, corresponding to various values of sample height, h. If these are plotted logarithmically, $\log dh/dt$ vs. $\log h$, as in Fig. 2, a straight line is obtained. This shows that under the test conditions employed, the relation between sample height and rate of compression is of the form:

$$\frac{dh}{dt} = kh^b, \tag{1}$$

where k and b are empirical constants. All the tests made on solders gave results that could be fitted with an equation of this type. At any one temperature the value of b was constant for a given solder, while the value of k varied with the load and sample volume but was independent of the initial sample height. A limited number of tests have indicated that all runs made on any one solder at a given temperature can be represented by an equation of the type:

$$\frac{dh}{dt} = K\frac{W^a h^b}{V^c}, \tag{2}$$

where W is the load, V the sample volume and K, a, b, and c are constants characteristic of the material.

It has been shown both theoretically [4] and experimentally [3, 5] that a cylindrical sample of a viscous material is compressed between parallel plates at a rate given by:

---

[4] O. Reynolds, *Phil. Trans., Lond.*, **177A**, 157 (1886).

[5] Ormandy, "The Engineer," **143**, 362, 393 (1927).

$$\frac{dh}{dt} = \frac{2\pi}{3\eta} \frac{Wh^5}{V^2}, \tag{3}$$

where η is the viscosity. This equation is of the form of Equation 2 with $b = 5$. In the solder tests, the values of b obtained were in general much larger than 5.0, so that these materials are not strictly viscous. On the other hand, the fact that the results are independent of the initial sample height indicates that the rate of flow is independent of the strain; while the fact that the curves of log dh/dt vs. log h continue to be linear at very low rates of flow indicates the absence of any yield point, or minimum stress required for flow. Hence the stress required for compression appears to be of the type which has been called quasi-viscous—wholly dependent on the velocity gradient.

Flow of this type, of which viscous flow is a special case, has been observed in colloidal solutions tested by the capillary tube method. In such solutions deWaele [6] found the rate of efflux proportional, not to the pressure (as in strictly viscous fluids [7]), but to a power of the pressure. Porter and Rao [8] have shown that this would be the case if it were assumed that the shearing stress (τ) is proportional to a power of the velocity gradient (dv/dx), or $\tau = \eta' \, (dv/dx)^{1/n}$, where η' and n are constants; dv/dx is the velocity gradient normal to the plane in which τ is the tangential stress. Assuming this relation, the case of compression between parallel plates has been shown [3] to be given by:

$$\frac{dh}{dt} = C \frac{W^n h^{\frac{5(n+1)}{2}}}{V^{\frac{3n+1}{2}}}, \tag{4}$$

where C is a constant inversely proportional to η'. The development of Equation 4 employs approximations corresponding to those used in obtaining Equation 3, involving the assumption that the diameter of the sample is large compared with its height.

As a matter of at least theoretical interest a series of runs were made on one solder (No. 2) employing different loads and sample volumes, and the data thus obtained were analyzed with reference to their agreement with Equation 4. This analysis is given in an appendix to this paper, and shows that all runs could be fitted without significant variation by an equation of the form of Equation 1, and that the values of b for these different runs did not differ significantly from one another. On the other hand, the values of k thus obtained did differ significantly from an expression of the form $k = K(W^a/V^c)$, in which K, a, and c were selected to give the best agreement with the data, indicating that Equation 2, as applied to this material, is not strictly correct. The values of a and c thus obtained differed considerably from the values of n and $(3n + 1)/2$ respectively that would be required if Equation 4 applied, n being given by equating b to $5(n + 1)/2$ in accordance with this equation. On the other hand, these differences were not indicated as necessarily significant, so that it appears that Equation 4 applies to these data to a fair approximation, the divergence being perhaps due wholly to the approximations employed in the theoretical development.

---

[6] A. de Waele, *J. Oil & Color Chem. Assn.*, **6**, 33 (1923).

[7] Bingham, "Fluidity and Plasticity," McGraw-Hill.

[8] Porter and Rao, *Trans. Faraday Soc.*, **23**, 311 (1927).

Regardless of the accuracy of Equation 4, the considerations discussed above indicate quite clearly that the flow of solders in the wiping range is quasi-viscous (independent of strain, and of a yield point requirement) and hence similar to the flow of many colloidal solutions. This conclusion is in agreement with the fact that at these temperatures solder consists of a solid phase dispersed in a liquid phase. It is furthermore apparent that the quantity $b - 5$ (or of $1 - n$, if use is made of the theory given) is a measure of the departure of the material from a strictly viscous condition.

Evaluation of Flow Constants

With the exception noted above, the runs made with the various solders tested were confined to a single sample size and a single load. In each case values of b were obtained by plotting log dh/dt vs. log h, values of the former quantity being found by plotting tangents as described. To illustrate the reproducibility of points in the plot of log rate vs. log height, Fig. 2 has been prepared from Fig. 1. Curves I and II of Fig. 1 were plotted to different scales of sample height and the tangents drawn. In Fig. 2 are plotted values of log dh/dt against log h from the two curves. It is seen that the line would be drawn in practically the same place if either set of points was taken alone.

In comparing solders it is evident that the sample height after a given time is a rough measure of relative consistency, the softer material showing the lower sample height. But from Equation 1 it is evident that such a comparison may be misleading for materials showing different values of b. This is brought out in Fig. 3, in which are given the height vs. time curves for two different solders. It will be seen that the curves cross each other, and that the consistency of such samples would, if rated on such a basis, depend upon the time selected for the comparison. In comparing two solders tested under the same conditions, consideration must be given to both constants of Equation 1 (k and b). For purposes of qualitative comparison an equivalent procedure is to consider the sample heights at two different times for each sample listed.

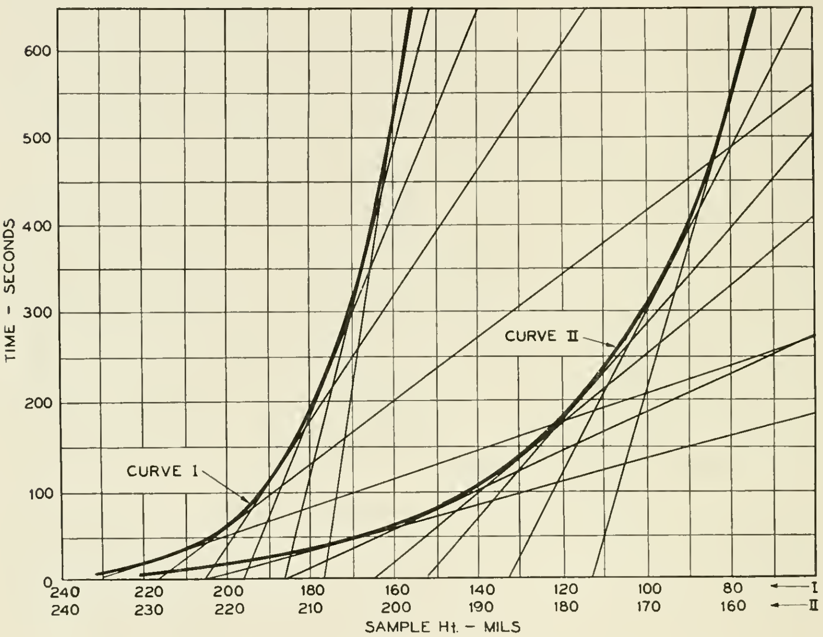


Fig. 1—Plot of a typical set of plasticity data. Solder No. 5—157° C.

Comparison of Plasticity with Solid Liquid Ratio

A summary of the results of the plasticity tests is given in Table II. In this are included values of the percentage of liquid present in the solder at the test temperature, calculated from the equilibrium diagrams by the method advocated by Tammann.[9]

Sample heights after 30 minutes of compression are plotted in Fig. 4 against per cent liquid phase present under the conditions of test. As one would expect, the points for any given solder can be placed on a smooth curve, in which an increase in the proportion of liquid phase

---

[9] Gustav Tammann, "A Textbook of Metallography."

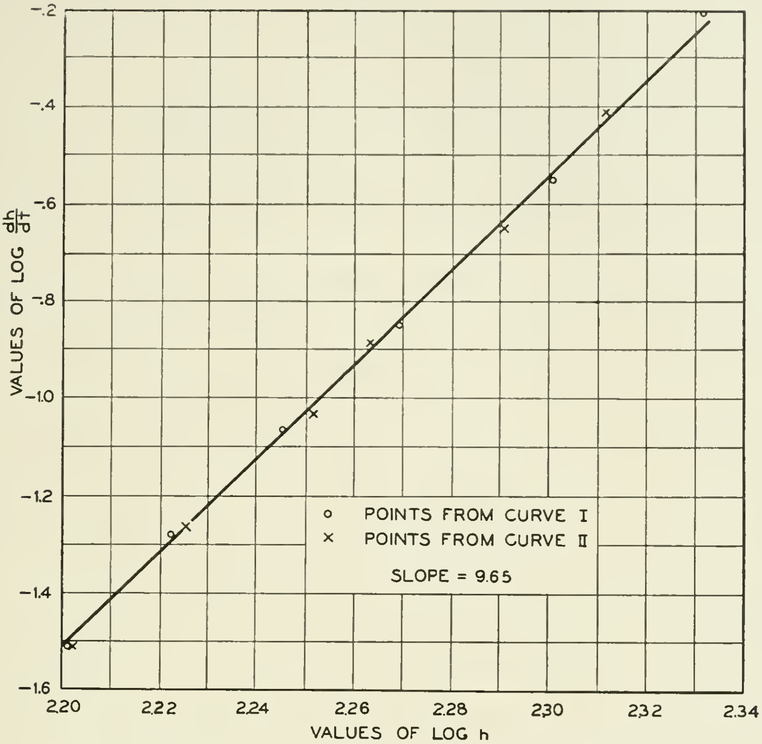


Fig. 2—Log rate vs. log height curve plotted from data of Fig. 1.

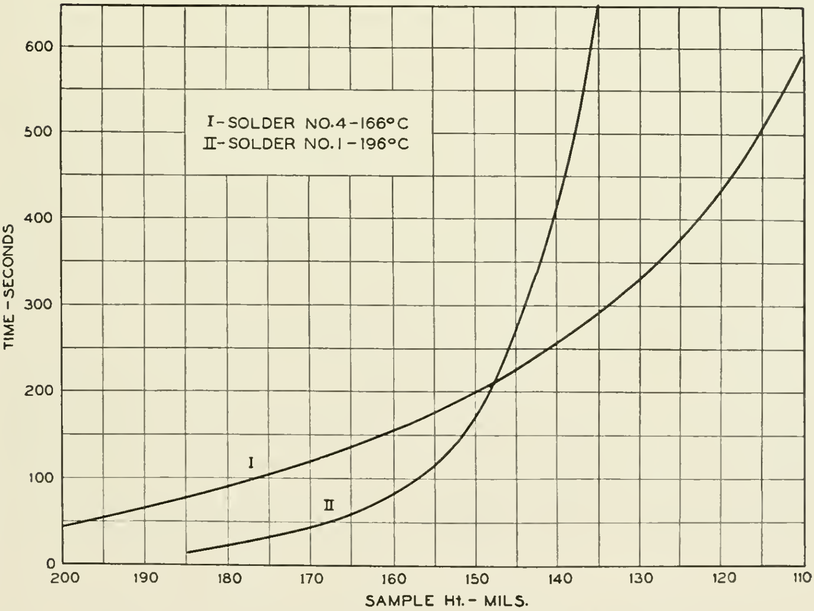


Fig. 3—Effect of type of flow upon sample height vs. time curves.

TABLE II

Results of Plasticity Tests

Solder	Temperature °C.	Sample Height 1 Min.	Sample Height 10 Min.	Sample Height 30 Min.	Per Cent Liquid *	b	Average b
No. 1	196	162	132		47.5	14.58	16.0
	204	145	125	116	51.7	19.12	
	210	117	104	94	55.1	14.20	
No. 2	183	150	128	118	52.4	15.80	16.2
	185	143	124	114	53.4	15.56	
	194	127	111	103	57.8	17.24	
No. 3	170	132	118	110	43.4	25.60	21.0
	184	130	114	105	51.2	18.14	
	192	115	101	93	55.8	19.30	
No. 4	162	266	245	203	55.5	3.88	3.73
	166	197	110	83	59.4	3.58	
	170	42	33	29	62.4	†	
No. 5	149		255			6.02	8.24
	153	245	198	166		6.39	
	157	200	158	140		9.65	
	161	185	154	127		10.90	
No. 6	140	262	244	224		12.3	13.5
	148	214	175	156		10.2	
	156	164	139	127		15.9	
	160	143	124	115		15.6	

* Equilibrium Diagrams used:
Solder No. 1 and 2—Rosenhain & Tucker.
Solder No. 3—Stoffel.
Solder No. 4—International Critical Tables, Vol. IV.
† Sample heights taken were too low to render the data significant.

corresponds to a decrease in the sample height at any specified time. However, the curves for the different solders do not fall together. As a matter of fact, no such agreement would be expected after studying results such as those plotted in Fig. 3. The curves given in this figure show that if a number of solders are tested at temperatures corresponding to a given solid-liquid ratio, they may fall in quite different orders with respect to each other in a comparison of sample heights, according to the time chosen for the comparison.

Still less agreement among the solders is shown when values of b are compared with percentage of liquid phase. There seems to be no relation here; in fact, for a given solder alteration of the temperature and hence of the proportion of liquid phase, does not seem to have much effect upon the value of b.

A survey of the curves and data shows that the possibility of drawing any conclusions from the solid-liquid ratios regarding the plasticity to be expected or regarding the quality of the solder must be dismissed.

Comparison of Workability and Temperature Gradient of Plasticity

In Fig. 5 are shown the plasticity-temperature gradients for the various solders. Here comparative values of plasticity for any one solder are taken as given by the sample heights after ten minutes of compression. It is seen that solders No. 1, 2 and 3 show good (low) gradients, with No. 3 seeming to be superior, while Nos. 5, 6 and especially No. 4, show a large variation in plasticity with temperature. Since, if it is to be properly worked, a wiping solder must not vary rapidly in consistency with temperature, this can be used as one criterion for choosing a satisfactory solder, and on this basis the solders tested should be placed in an order consistent with Fig. 5.

Such a rating is quite in agreement with actual working characteristics, which rank solder No. 3 first, No. 4 last, the others falling in between in essentially the order which Fig. 5 would indicate.

It is now of interest to determine if there exists any relation between the rapidity with which the plasticity varies with temperature and the values of b, the slope of the log rate vs. log height curve. Reference to Table II will show that there is a wide divergence in the values of b found for the different solders. It is further seen that the difference exhibited from one alloy to the next is greater than the difference between individual runs on one solder at different temperatures. This virtual constancy of b for any one solder justifies, it may be noted in passing, the use of sample heights after fixed intervals of compression as comparative measures of plasticity, as in Figs. 4 and 5. There would be no such justification for using such sample heights in comparing different alloys having different values of b.

If then b is regarded as a constant for any one solder, independent of temperature, mean values of b for each solder may be computed from the values obtained at different temperatures. Such mean values have been computed, and are included in Table II. Comparison of these with the curves of Fig. 5 shows a striking agreement between the order of the solders based on increasing temperature gradients and that based on decreasing values of b. This suggests that the temperature gradient of plasticity is the lower, the more removed is the character of the flow from that of a strictly viscous liquid. Furthermore, it is apparent that if the value of b is an inverse measure of the temperature gradient of plasticity, and if the latter is in turn an inverse measure of the workability of a solder, then the value of b is a direct measure of workability.

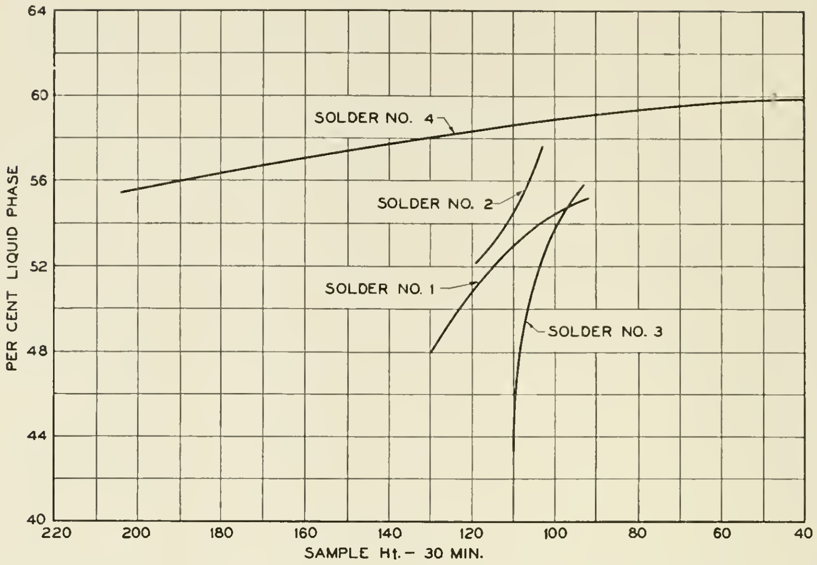


Fig. 4—Relation of plasticity to per cent liquid phase in four of the solders tested.

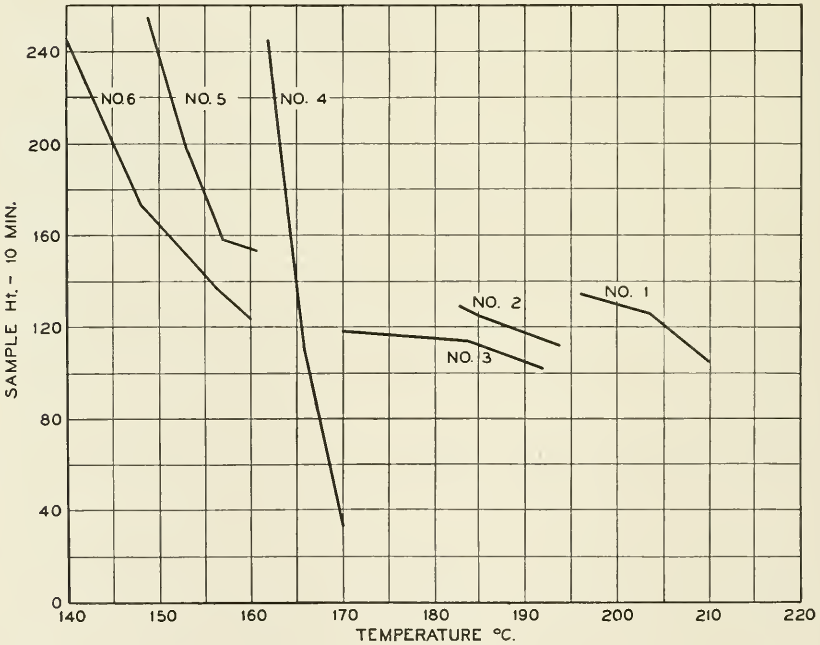


Fig. 5—Variation of plasticity with temperature, showing the superiority of solders Nos. 1, 2 and 3.

As the value of *b* can be determined in a single test with the plastometer, its relation to workability should permit of the rapid evaluation of the latter property, provided the relation holds in general. Applying this criterion to the solders tested, No. 4, showing virtually a viscous type of flow, ($b = 5$), would be at once thrown out. Solder No. 3 would be classed as a superior solder from the standpoint of workability. Nos. 1 and 2 could be classed as satisfactory. The use of Nos. 5 and 6 would be questionable, with No. 6 having the best chance for success. Practical tests on these last two solders are insufficient to completely confirm this classification, but general indications are that it is correct. Practice has completely confirmed the classification of the other four.

Usefulness of Plastometer in Research and Control Testing

It is believed that the work which has been done on these six solders has furnished information which justifies the use of the plastometer in the future when any group of alloys is to be investigated as to working properties. It appears that the principal demand upon a workable solder is a wide temperature range in which it can be worked. Whether or not a solder is suitable in this respect can very well be determined by the plastometer either by making runs at several temperatures or by determining *b* at some temperature representative of the wiping range.

A further possible use of the plastometer is in testing solders manufactured or purchased subject to requirements as to their composition. As a rule, the sample height after a given time of compression is quite different for alloys differing in composition. This is illustrated by the curves of Fig. 6, which represent the data obtained with the plastometer on a number of alloys. These show that such a test of composition would not be infallible in a wide application, as the quite dissimilar alloys *C* and *F* give quite similar curves. In distinguishing an excess or deficiency of one component in any given series of alloys, however, the test should be quite sensitive.

If the plastometer were used for this purpose, the allowable limits of composition would correspond to limits of sample height in satisfactory samples. Consider, for example, the case of a lead-tin solder of a composition between that of solders Nos. 1 and 2. A convenient temperature of test would be 195° C. At this temperature, the sample height after ten minutes would be 110 mils for a solder of composition 65 per cent Pb, 35 per cent Sn, and 137 mils for one of composition 60 per cent Pb, 40 per cent Sn. Since the plastometer results are reproducible within 2 mils, the accuracy in detecting variation in composition should be easily within 0.5 per cent. Whether the possible usefulness of the test could be realized economically in engineering practice has not as yet been ascertained.

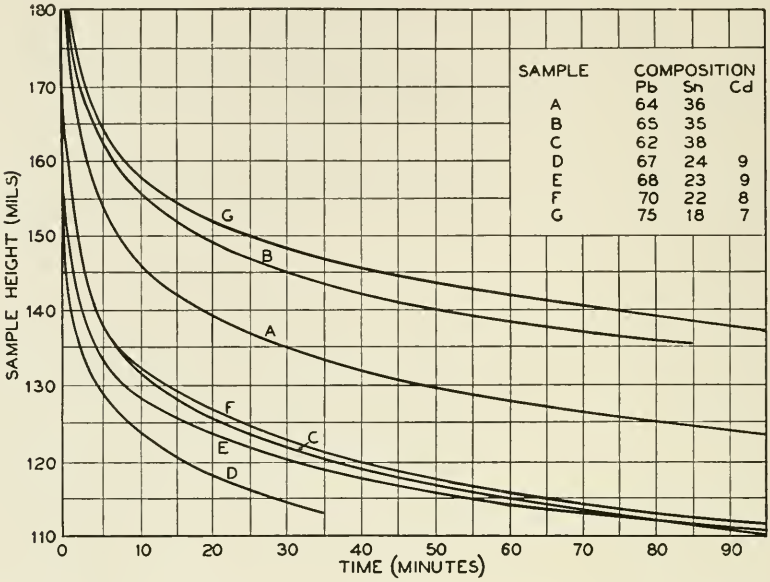


Fig. 6—A group of plasticity curves illustrating the effect of composition upon plasticity.

Segregation

Experimental Determination

The plastic deformation of a material composed of solid and liquid phases, such as a solder during wiping, presents a condition favorable to segregation, and this phenomenon has often been thought to play an important role in determining the quality of a solder. Especially has this been true in regard to porosity. Now, if a sample of solder which has a tendency to segregate is subjected to a plasticity test, a considerable amount of liquid phase should be pushed to the outer edges of the sample during the first part of the run. Since the liquid phase always contains that part of the sample which later solidifies as the eutectic, it follows that after the sample has been allowed to cool, the outer edges should be richer in eutectic than the central portion.

This, as a matter of fact, was found to be true, and measurements of segregation were made upon a number of samples taken from the plastometer. These samples were usually about 5 cm. in diameter. They were cut diametrically, polished and etched.

Photographs were taken with a magnification of 25 ×, which allowed 0.5 cm. of the sample to be photographed at one time. Three photographs were taken along the radius of each sample, one bordered by the axis of the sample, one by the edge, and one mid-way between these two.

These photomicrographs were placed in a beam of light of constant intensity. Since the eutectic appears as a light background, while the particles which separate out as solid above the eutectic temperature appear dark, the light transmitted through the plate is a comparative measure of the relative amount of eutectic present in different parts of the sample. The variation in light transmitted as the picture is passed before the beam therefore indicates the amount of segregation. The transmitted light was directed on a cæsium photo-cell and the amplified photo-electric current measured, four measurements being required to span each plate, making twelve measurements on each sample.

Table III gives the results of these measurements. In this table, only the averages of the four measurements of each photomicrograph are given.

TABLE III

Segregation Measurements

Light Transmitted Measured in Milliamperes of Amplified Photo-electric Current

Sample		Ligh Transmitted—Ma.		
Solder	Temp. ° C.	.25 cn . from Center	1.25 cm. from Center	2.25 cm. from Center
1.............	204	0.36	1.08	2.00
	210	0.35	1.07	1.55
2.............	183	0.88	1.58	3.20
	186	0.44	0.69	1.16
	194	0.22	0.69	0.81
3.............	170	0.45	0.57	1.06
	184	0.83	0.87	0.99
	194	0.49	0.57	0.69
4.............	162	0.63	0.49	0.58
	166	0.66	0.63	0.63
	170	0.61	0.60	0.65

Results—Comparison with Porosity

The only conclusion that can be drawn from this tabulation of results is that segregation is not an important factor in porosity. Among solders Nos. 1, 2 and 3, good and poor solders alike show segregation to a large extent. No marked difference seems to exist, for example, between the segregation of solder No. 2, which forms non-porous joints, and solder No. 1, which is a poor solder from the standpoint of porosity. On the other hand, in solder No. 4, an extremely poor solder, segregation is so slight as to be negligible. In Fig. 7 are plotted the data for solders Nos. 1 and 4, the former showing considerable segregation and the latter practically no segregation of solid from liquid phase.

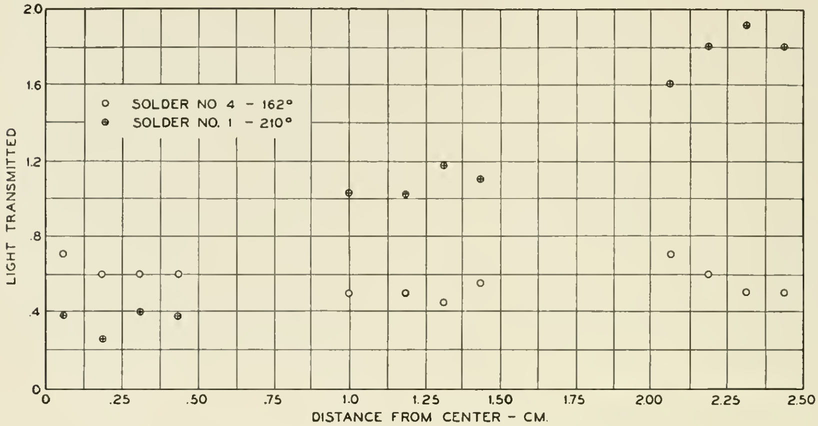


Fig. 7—Segregation measurements on two solder samples.

It is interesting to note how well the segregation measurements confirm the plasticity data. Solders Nos. 1, 2, and 3, showing clearly the quasi-viscous type of flow, also show marked segregation, revealing the reluctance of the solid particles to move and their consequent piling up near the center of the sample. In solder No. 4, the flow is of the viscous type, and consequently there is no segregation. The solid phase and the liquid phase do not separate, but flow together as a viscous whole.

PARTICLE SIZE

Relation to Porosity

It was thought that in solders forming porous joints the solid phase may, at wiping temperatures, be present as particles relatively large in size, in which case the liquid must be in larger recesses than would be the case if the particles were small. The cohesion of the solder while it is being worked and while it is solidifying must depend to some extent upon the adhesion between solid and liquid. The larger the particles, the greater the distances through which the surface forces must operate in hindering the free motion of the liquid phase. The liquid phase in a two-phase system where the particles are large should then be only loosely bound to the mass, and though there might be no gross tendency toward segregation, this liquid phase should in working be squeezed out, leading to cavity formation and consequent porosity.

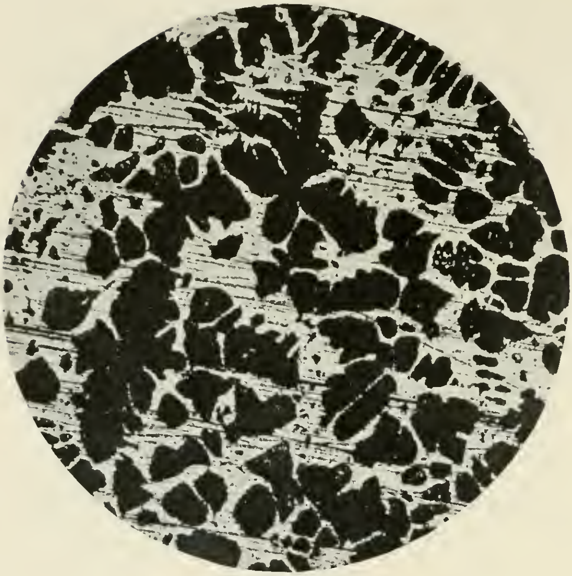
Experimental

To test this explanation of porosity two samples of solder No. 1 were quenched from 210° C. and two of solder No. 2 from 200° C., at which temperatures these solders exhibit practically the same plasticity. Though the data are too meagre to allow generalizations to be made, the results in both sets of samples were well in line with the theory. Microscopic examination showed a much coarser particle structure in solder No. 1 than in solder No. 2, as illustrated by Fig. 8. In this instance at least the solder which forms porous joints tended to form solid particles at wiping temperatures larger than those formed in the solder which is acceptable from the standpoint of porosity.

Particle Size and Porosity as Related to Wettability

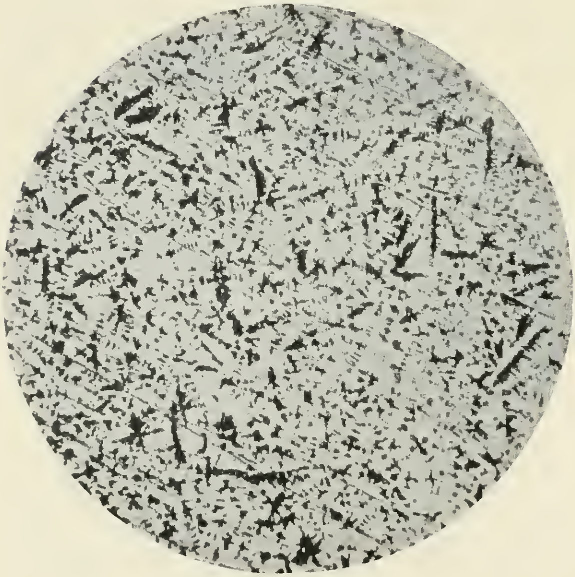
It appears that this test may have a significance beyond that just attached to it, and since it may conceivably be used again in the study of solders, it is perhaps well to point out a further explanation based on the theory of interfacial tension.

As was stated earlier in this paper, one demand upon a wiping solder is that it shall wet the cable covering readily. This is necessary if porosity is to be avoided, since areas of imperfect wetting will form tracks between the solder and the lead allowing passage of air or liquid along the interface. Further, the liquid phase of the solder must wet the solid phase of the solder; otherwise passages will exist in the body of the joint. There is some question as to whether porosity in a joint is more commonly due to lack of adherence between the sheath and the solder or whether complete air passages may exist throughout the solder itself. At any rate, there is no assurance that either type of failure does not exist. These two types of porosity are, however, very similar phenomena from the standpoint of wettability, since both the cable sheath and the solid particles of the solder are composed chiefly of lead. If the liquid phase of the solder will wet the sheath, it will also wet the solid particles of lead within the solder, and neither type of failure will be likely to occur. The problem resolves itself into one of the wettability of lead by the liquid phase of the alloy. Therefore, if it is possible to determine, at equal plasticities, the relative



POROUS
WIPING SOLDER
65% Pb. – 35% Sn.
AIR COOLED TO 210° C.
AND QUENCHED

LARGE
CRYSTALS



NON — POROUS
WIPING SOLDER
60% Pb. – 40% Sn.
AIR COOLED TO 200° C.
AND QUENCHED

SMALL
CRYSTALS

Fig. 8—The porosity of wiping solders as related to the crystal size of samples air cooled to equivalent consistencies and quenched.

adherence of liquid to solid for different solders, the relative likelihood of porous joints occurring should be thereby indicated.

Wettability depends upon interfacial tension. As pointed out by Osterhof and Bartell,[10] high wetting power of a liquid for a solid means a low interfacial tension between the two, which in turn means a high adhesion tension.

Furthermore, if the interfacial tension between a liquid and a dispersed solid is high, the formation of large grains, crystals, or flocculates will take place, thus diminishing the amount of solid-liquid interface and the total free energy of the system. This has been used to account for the increased solubility of small particles in contact with a saturated solution, with the consequent growth of large crystals [11, 12] and in paint technology to explain the formation of groups of particles or flocculates.[13, 14] If, on the other hand, the interfacial tension is low, as for solders of high wettability, the particles have a tendency to remain small. This fact has been used to account for the high development of surface in colloidal solutions.

This would lead to the conclusions that in solder No. 1, where large solid particles exist, the liquid phase wets the lead dispersed within it and hence also the lead of the cable sheath but poorly, while in solder No. 2, exhibiting finer grain formation, the wetting power of the liquid phase for lead is high. These considerations alone would point to a conclusion which has been shown to be the case—that solder No. 1 is more likely to form porous joints than solder No. 2.

SUMMARY

An attempt has been made to correlate important qualities of wiping solders with measurements of plasticity, segregation, and particle size.

A study of six solders has shown that their deformation in the working range is in accord with the theory of quasi-viscous flow, which assumes the shearing stress to be proportional to a fractional power of the velocity gradient.

It appears that what is required of a solder, in order that it can be properly worked, is a low variation in plasticity with temperature. This allows the solder to be worked for a comparatively long period while it is losing heat to the atmosphere. This requirement appears to be satisfied for solders whose resistance to deformation differs most markedly in character from that of a viscous liquid.

---

[10] H. J. Osterhof and F. E. Bartell, *J. Phys. Chem.*, **34**, 7 (1930).

[11] H. Freundlich, "Colloid and Capillary Chemistry," E. P. Dutton (1922).

[12] Willows and Hatschek, "Surface Tension and Surface Energy," P. Blakiston & Son (1919).

[13] Bartell and Van Loo, *Ind. & Eng. Chem.*, **17**, 1051 (1925).

[14] Wm. Green, *Ind. & Eng. Chem.*, **15**, 122 (1923).

Determinations of the rate of compression of solders at wiping temperatures between parallel plates have been found to conform to the equation:

$$\frac{dh}{dt} = kh^b,$$

in which dh/dt is the rate of compression, h the sample height, and k and b are constants, the latter depending only on the character of the material. For viscous materials $b = 5.0$; the larger the value of b, apparently, the more does the flow differ from that of a viscous liquid. Solders having large values of b (15–25) have low temperature gradients of plasticity and good workability, in agreement with the conclusions noted above.

It has been shown that the plasticity cannot be predicted from the solid-liquid ratio.

Evidence is presented that segregation is not responsible for the porosity of wiped joints, and that this defect may be related to the particle size of the solid phase at wiping temperatures. The relation of particle size to cavity formation and to the cohesion and wetting power of the solder is discussed.

APPENDIX

As stated in the body of this paper, a series of runs were made on solder No. 2 at 183° C. to determine how closely the results of such tests agree with the theoretical expression for quasi-viscous flow given as Equation (4) above. Only five test runs were made, the minimum sufficient to indicate the existence of such agreement. The sample volumes and loads used in these runs are listed in Table IV. For each run there was prepared a plot of sample height (h) vs. time (t), and values of dh/dt were determined from the slopes of tangents drawn as

TABLE IV

Run No.	V c.c.	W kgs.	b	σ_b	log k	S	log k'	S'	Δ log k'
1	5.45	13.6	16.10	4.32	−34.36	0.0384	−34.472	0.0352	0.073
2	7.34	13.6	17.73	3.99	−37.58	0.0351	−35.556	0.0473	0.122
3	11.01	13.6	15.68	1.85	−35.53	0.0225	−36.577	0.0257	0.067
4	7.30	20.8	15.76	1.43	−34.38	0.0131	−35.209	0.01553	0.098
5	7.33	27.2	15.51	2.83	−33.51	0.0249	−34.853	0.0286	0.078

$$\bar{b} = 16.156$$
$$\sigma_{\iota}' = 0.906$$
$$a = 1.656$$
$$c = 6.875$$
$$\log K = -31.359$$

shown in Fig. 1. There were thus obtained for each run a set of six or seven pairs of values of log dh/dt vs. log h.

The first question to be determined is whether for any run these pairs of values agree with a relation of the form of Equation 1, which can be written:

$$\log \frac{dh}{dt} = \log k + b \log h. \tag{5}$$

As sample height readings appear to be reproducible in such tests to within about one per cent, the standard deviation of values of log h about their correct values may be estimated as rather less than 0.004 (log 1.01). The time readings being more accurate, values of log dh/dt should vary with a standard deviation of like magnitude. If the data can be fitted by an equation of the form of Equation 5, estimates of k and b can be obtained by minimizing the sum of the squares of the deviations of the experimental points from the straight line, these deviations being measured along normals to the line. As for these data b is large, these normal deviations are nearly equal to their horizontal components, the deviations of the observed from the calculated values of log h, and hence the sum of the squares of the latter were minimized. Writing X_i for values of log h and Y_i for corresponding values of log dh/dt, the following computations were carried out for each run (n being the number of pairs of values employed in each case):

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \qquad \overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n},$$

$$\sigma_x^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}, \qquad \sigma_y^2 = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{n},$$

$$b = \frac{n\,\sigma_y^2}{\sum_{i=1}^{n} X_i\,Y_i - n\,\overline{X}\,\overline{Y}}, \qquad \log k = \overline{Y} - b\,\overline{X},$$

$$S^2 = \frac{\sum_{i=1}^{n} (Y_i - \log k - b\,X_i)^2}{n-2}.$$

Values of b, log k, and S thus determined are included in Table IV. The quantity S is an estimate of the standard deviation of the observed from the calculated values of log dh/dt evaluated as recommended by Fisher,[15] page 118. To determine in each case whether these deviations are significant, they may be compared with those that might be expected on the basis of the estimated precision of measurement. As the value of b is approximately 16 in each case, an error of 0.004 in log h corresponds to an error of 0.064 in log dh/dt. The values of S computed are all smaller than this amount, indicating that the differences between observed values of log dh/dt and those calculated from Equation 1 are due to experimental variations.

The next question to be determined is whether the differences observed in the values of b given by the different runs are significant. In Table IV is included the mean value $\bar{b}$ of the values of b, together with the estimated standard deviation of the individual values of b from their true mean, taken as given by:

$$\sigma_b' = \sqrt{\frac{\sum(b - \bar{b})^2}{n - 1}}.$$

This estimate of the standard deviation of the actual values of b may be compared with estimates of the standard deviation of b made for each run on the basis of the variability of the data. An expression for such an estimate is given by Fisher (loc. cit.) as:

$$\sigma_b{}^2 = \frac{nS^2}{\sigma_x{}^2}.$$

Values of σ_b as given by this last expression are included in Table IV. These are larger than the value (0.906) of σ_b', computed as described, and it is therefore evident that the data are consistent with the hypothesis that b is constant throughout the series of runs.

To obtain fairer estimates of log k for the subsequent computation, the mean value of b, $\bar{b}$, was taken as giving the slope of the straight line of Equation 5. On this basis the values of log k are given by:

$$\log k' = \bar{Y} - \bar{b}\bar{X}.$$

While the values of S are given by:

$$S'^2 = \frac{\sum_{i=1}^{n} (Y_i - \log k' - \bar{b}X_i)^2}{n - 1}.$$

Values of log k' and S' are given in Table IV. The next step in the computation is to determine if the values of log k thus determined vary significantly from those given by Equation 2 above, or by:

---

[15] R. A. Fisher, "Statistical Methods for Research Workers," London (1925).

$$\log k' = \log K + a \log W - c \log V. \qquad (6)$$

For this purpose there were determined values of log K, a and c for which the sum of the squares of the differences of observed and calculated values of log k' are a minimum, the components of deviation due to variations in values of log W and log V being neglected, as these are known to a higher order of accuracy than are values of log k'. Following the usual procedure for calculating partial regression coefficients, the following computations were performed, writing Z_i, Y_i, and X_i for corresponding values of log k', log W and log V:

$$\bar{Z} = \frac{\sum_{i=1}^{n} Z_i}{n}, \qquad \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}, \qquad \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n},$$

$$\sigma_Z^2 = \frac{\sum_{i=1}^{n} (Z_i - \bar{Z})^2}{n}, \quad \sigma_Y^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n}, \quad \sigma_X^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n},$$

$$\sum xy = \sum_{i=1}^{n} X_iY_i - n\bar{X}\bar{Y}, \qquad \sum yz = \sum_{i=1}^{n} Y_iZ_i - n\bar{Y}\bar{Z},$$

$$\sum zx = \sum_{i=1}^{n} Z_iX_i - n\bar{Z}\bar{X},$$

$$a = \frac{n\sigma_x^2\sum yz - \sum xy \cdot \sum xz}{n^2\sigma_x^2\sigma_y^2 - (\sum xy)^2}, \qquad -c = \frac{n\sigma_y^2\sum xz - \sum xy \cdot \sum yz}{n^2\sigma_x^2\sigma_y^2 - (\sum xy)^2},$$

$$\log K = \bar{Z} - a\bar{Y} + c\bar{X}.$$

The values of a, c, and log K thus obtained are included in Table IV, together with values headed Δ log k', which are the differences between the values of log k' listed in Table IV and those calculated from Equation 6, using the listed values a, c, and log K. To determine whether these differences are significant or not, the probability of obtaining a value in error to that extent may be calculated for each run from the value (S') of the standard deviation of log k' estimated by the method given by Fisher (loc. cit.). These probabilities are evaluated from Table IV of the text cited, taking n as $n' - 3$, where n' is the number of pairs of values of log dh/dt and log h used to evaluate log k', and taking t as:

$$t = \frac{\Delta \log k'\sqrt{n'}}{S'}.$$

The values of t thus computed range from 5.4 to 16.7 and correspond in every case to a probability less than 0.01, so that not one of the actual values of log k' should be experimentally in error to an extent sufficient to account for its deviation from the value given by Equation 6. Hence Equation 2, is at best only an approximation to a correct expression for dh/dt.

Of course, Equation 2 is actually quite a fair approximation to an expression for dh/dt, as is evident in comparing the differences Δ log k' with the differences among the values of log k' themselves. In fact, if values of log k' are plotted against log W for a constant value of V, or against log V for a constant value of W, quite good straight lines are obtained, and the inaccuracy of the approximation can only be brought out by an analysis such as is here given. Whether this inaccuracy is due to the approximations employed in the theoretical development of Equation 4, or to an error in the basic assumption that $\tau = [dv/dx]^{1/n}$, cannot, of course, be determined. It is, however, of interest to see if the values of a and c are related to that of b in the manner required by Equation 4. For $b = 16.156$ (the value of $\bar{b}$ observed), Equation 4 requires that $a = 5.462$ and $c = 8.693$. The observed values of a and c are 1.656 and 6.875 respectively. It remains to determine if these differences may be accidental, provided the values of log k' differ from those given by Equation 6 as greatly as observed. Following Fisher (page 133), the probability of a divergence between observed and calculated values of c of $c_1 - c_2$ is given by entering Table IV of the text with $n = n' - 3$, where n' is the number of cases (5), and with t given by:

$$t = \frac{C_1 - C_2}{\sqrt{\dfrac{[Z - (\bar{Z} + a\bar{Y} - c\bar{X})]^2}{n-3} \cdot \dfrac{n\sigma_y{}^2}{n^2\sigma_x{}^2\sigma_y{}^2 - (\sum xy)^2}}} = 3.58.$$

This corresponds to a probability between 0.05 and 0.10, which indicates that the divergence may be due to chance, though it is more likely to represent a real difference. Similarly the probability of obtaining the observed difference $a_1 - a_2$ is given from the table by entering with $n = 2$ and with t given by:

$$t = \frac{a_1 - a_2}{\sqrt{\dfrac{[Z - (\bar{Z} + a\bar{Y} - c\bar{X})]^2}{n-3} \cdot \dfrac{n\sigma_x{}^2}{n^2\sigma_x{}^2\sigma_y{}^2 - (\sum xy)^2}}} = 5.85.$$

This corresponds to a probability between 0.02 and 0.05, which suggests more strongly than does the other case that the divergence is real. For so few observations, however, the distribution theory on which Fisher's method is based cannot give a definite indication when the probabilities lie so close to the arbitrary dividing line (0.05) between likely and unlikely results. The analysis therefore leaves open the question as to whether the theoretical relation given by Equation 4 between a, b and c is valid within the limits of variability imposed by the approximate nature of any relation of the form of Equation 2. That any equation of this form is only approximately in agreement with these data is conclusively shown by the above analysis.

Regeneration Theory

By H. NYQUIST

Regeneration or feed-back is of considerable importance in many applications of vacuum tubes. The most obvious example is that of vacuum tube oscillators, where the feed-back is carried beyond the singing point. Another application is the 21-circuit test of balance, in which the current due to the unbalance between two impedances is fed back, the gain being increased until singing occurs. Still other applications are cases where portions of the output current of amplifiers are fed back to the input either unintentionally or by design. For the purpose of investigating the stability of such devices they may be looked on as amplifiers whose output is connected to the input through a transducer. This paper deals with the theory of stability of such systems.

PRELIMINARY DISCUSSION

WHEN the output of an amplifier is connected to the input through a transducer the resulting combination may be either stable or unstable. The circuit will be said to be stable when an impressed small disturbance, which itself dies out, results in a response which dies out. It will be said to be unstable when such a disturbance results in a response which goes on indefinitely, either staying at a relatively small value or increasing until it is limited by the non-linearity of the amplifier. When thus limited, the disturbance does not grow further. The net gain of the round trip circuit is then zero. Otherwise stated, the more the response increases the more does the non-linearity decrease the gain until at the point of operation the gain of the amplifier is just equal to the loss in the feed-back admittance. An oscillator under these conditions would ordinarily be called stable but it will simplify the present paper to use the definitions above and call it unstable. Now, this fact as to equality of gain and loss appears to be an accident connected with the non-linearity of the circuit and far from throwing light on the conditions for stability actually diverts attention from the essential facts. In the present discussion this difficulty will be avoided by the use of a strictly linear amplifier, which implies an amplifier of unlimited power carrying capacity. The attention will then be centered on whether an initial impulse dies out or results in a runaway condition. If a runaway condition takes place in such an amplifier, it follows that a non-linear amplifier having the same gain for small current and decreasing gain with increasing current will be unstable as well.

STEADY-STATE THEORIES AND EXPERIENCE

First, a discussion will be made of certain steady-state theories; and reasons why they are unsatisfactory will be pointed out. The most obvious method may be referred to as the series treatment. Let the complex quantity $AJ(i\omega)$ represent the ratio by which the amplifier and feed-back circuit modify the current in one round trip, that is, let the magnitude of AJ represent the ratio numerically and let the angle of AJ represent the phase shift. It will be convenient to refer to AJ as an admittance, although it does not have the dimensions of the quantity usually so called. Let the current

$$I_0 = \cos \omega t = \text{real part of } e^{i\omega t} \tag{a}$$

be impressed on the circuit. The first round trip is then represented by

$$I_1 = \text{real part of } AJe^{i\omega t} \tag{b}$$

and the nth by

$$I_m = \text{real part of } A^nJ^ne^{i\omega t}. \tag{c}$$

The total current of the original impressed current and the first n round trips is

$$I_n = \text{real part of } (1 + AJ + A^2J^2 + \cdots A^nJ^n)e^{i\omega t}. \tag{d}$$

If the expression in parentheses converges as n increases indefinitely, the conclusion is that the total current equals the limit of (d) as n increases indefinitely. Now

$$1 + AJ + \cdots A^nJ^n = \frac{1 - A^{n+1}J^{n+1}}{1 - AJ}. \tag{e}$$

If $|AJ| < 1$ this converges to $1/(1 - AJ)$ which leads to an answer which accords with experiment. When $|AJ| > 1$ an examination of the numerator in (e) shows that the expression does not converge but can be made as great as desired by taking n sufficiently large. The most obvious conclusion is that when $|AJ| > 1$ for some frequency there is a runaway condition. This disagrees with experiment, for instance, in the case where AJ is a negative quantity numerically greater than one. The next suggestion is to assume that somehow the expression $1/(1 - AJ)$ may be used instead of the limit of (e). This, however, in addition to being arbitrary, disagrees with experimental results in the case where AJ is positive and greater than 1, where the expression $1/(1 - AJ)$ leads to a finite current but where experiment indicates an unstable condition.

The fundamental difficulty with this method can be made apparent by considering the nature of the current expressed by (*a*) above. Does the expression cos ωt indicate a current which has been going on for all time or was the current zero up to a certain time and cos ωt thereafter? In the former case we introduce infinities into our expressions and make the equations invalid; in the latter case there will be transients or building-up processes whose importance may increase as n increases but which are tacitly neglected in equations (*b*) – (*e*). Briefly then, the difficulty with this method is that it neglects the building-up processes.

Another method is as follows: Let the voltage (or current) at any point be made up of two components

$$V = V_1 + V_2, \tag{f}$$

where V is the total voltage, V_1 is the part due directly to the impressed voltage, that is to say, without the feed-back, and V_2 is the component due to feed-back alone. We have

$$V_2 = AJV. \tag{g}$$

Eliminating V_2 between (*f*) and (*g*)

$$V = V_1/(1 - AJ). \tag{h}$$

This result agrees with experiment when $|AJ| < 1$ but does not generally agree when AJ is positive and greater than unity. The difficulty with this method is that it does not investigate whether or not a steady state exists. It simply assumes tacitly that a steady state exists and if so it gives the correct value. When a steady state does not exist this method yields no information, nor does it give any information as to whether or not a steady state exists, which is the important point.

The experimental facts do not appear to have been formulated precisely but appear to be well known to those working with these circuits. They may be stated loosely as follows: There is an unstable condition whenever there is at least one frequency for which AJ is positive and greater than unity. On the other hand, when AJ is negative it may be very much greater than unity and the condition is nevertheless stable. There are instances of $|AJ|$ being about 100 without the conditions being unstable. This, as will appear, accords closely with the rule deduced below.

NOTATION AND RESTRICTIONS

The following notation will be used in connection with integrals:

$$\int_I \phi(z)dz = \lim_{M\to\infty}\int_{-iM}^{+iM}\phi(z)dz, \tag{1}$$

the path of integration being along the imaginary axis (see equation 9), i.e., the straight line joining $-iM$ and $+iM$;

$$\int_{s+} \phi(z)dz = \lim_{M\to\infty}\int_{-iM}^{iM}\phi(z)dz, \tag{2}$$

the path of integration being along a semicircle[1] having the origin for center and passing through the points $-iM$, M, iM;

$$\int_C \phi(z)dz = \lim_{M\to\infty}\int_{-iM}^{-iM}\phi(z)dz, \tag{3}$$

the path of integration being first along the semicircle referred to and then along a straight line from iM to $-iM$. Referring to Fig. 1 it will be seen that

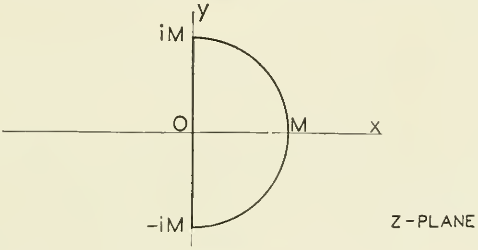


Fig. 1—Paths of integration in the z-plane.

$$\int_{s+} - \int_I = \int_C . \tag{4}$$

The total feed-back circuit is made up of an amplifier in tandem with a network. The amplifier is characterized by the amplifying ratio A which is independent of frequency. The network is characterized by the ratio $J(i\omega)$ which is a function of frequency but does not depend on the gain. The total effect of the amplifier and the network is to multiply the wave by the ratio $AJ(i\omega)$. An alternative way of characterizing the amplifier and network is to say that the amplifier is characterized by the amplifying factor A which is independent of time, and the network by the real function $G(t)$ which is the response caused by a unit impulse applied at time $t = 0$. The combined effect of the amplifier and network is to convert a unit impulse to the function $AG(t)$. Both these characterizations will be used.

---

[1] For physical interpretation of paths of integration for which $x > 0$ reference is made to a paper by J. R. Carson, "Notes on the Heaviside Operational Calculus," *B. S. T. J.*, Jan. 1930. For purposes of the present discussion the semicircle is preferable to the path there discussed.

The restrictions which are imposed on the functions in order that the subsequent reasoning may be valid will now be stated. There is no restriction on A other than that it should be real and independent of time and frequency. In stating the restrictions on the network it is convenient to begin with the expression G. They are

$$G(t) \text{ has bounded variation}, \quad -\infty < t < \infty. \tag{AI}$$

$$G(t) = 0, \quad -\infty < t < 0. \tag{AII}$$

$$\int_{-\infty}^{\infty} |G(t)|\, dt \text{ exists.} \tag{AIII}$$

It may be shown [2] that under these conditions $G(t)$ may be expressed by the equation

$$G(t) = \frac{1}{2\pi i} \int_I J(i\omega) e^{i\omega t} d(i\omega), \tag{5}$$

where

$$J(i\omega) = \int_{-\infty}^{\infty} G(t) e^{-i\omega t} dt. \tag{6}$$

These expressions may be taken to define J. The function may, however, be obtained directly from computations or measurements; in the latter case the function is not defined for negative values of ω. It must be defined as follows to be consistent with the definition in (6):

$$J(-i\omega) = \text{complex conjugate of } J(i\omega). \tag{7}$$

While the final results will be expressed in terms of $AJ(i\omega)$ it will be convenient for the purpose of the intervening mathematics to define an auxiliary and closely related function

$$w(z) = \frac{1}{2\pi i} \int_I \frac{AJ(i\omega)}{i\omega - z} d(i\omega), \qquad 0 < x < \infty, \tag{8}$$

where

$$z = x + iy \tag{9}$$

and where x and y are real. Further, we shall define

$$w(iy) = \lim_{x \to 0} w(z). \tag{10}$$

---

[2] See Appendix II for fuller discussion.

The function will not be defined for $x < 0$ nor for $|z| = \infty$. As defined it is analytic[3] for $0 < x < \infty$ and at least continuous for $x = 0$.

The following restrictions on the network may be deduced:

$$\lim_{y \to \infty} y |J(iy)| \text{ exists.} \tag{BI}$$

$$J(iy) \text{ is continuous.} \tag{BII}$$

$$w(iy) = AJ(iy). \tag{BIII}$$

Equation (5) may now be written

$$AG(t) = \frac{1}{2\pi i} \int_I w(z) e^{zt} dz = \frac{1}{2\pi i} \int_{s+} w(z) e^{zt} dz. \tag{11}$$

From a physical standpoint these restrictions are not of consequence. Any network made up of positive resistances, conductances, inductances, and capacitances meets them. Restriction (AII) says that the response must not precede the cause and is obviously fulfilled physically. Restriction (AIII) is fulfilled if the response dies out at least exponentially, which is also assured. Restriction (AI) says that the transmission must fall off with frequency. Physically there are always enough distributed constants present to insure this. This effect will be illustrated in example 8 below. Every physical network falls off in transmission sooner or later and it is ample for our purposes if it begins to fall off, say, at optical frequencies. We may say then that the reasoning applies to all linear networks which occur in nature. It also applies to other linear networks which are not physically producible but which may be specified mathematically. See example 7 below.

A temporary wave $f_0(t)$ is to be introduced into the system and an investigation will be made of whether the resultant disturbance in the system dies out. It has associated with it a function $F(z)$ defined by

$$f_0(t) = \frac{1}{2\pi i} \int_I F(z) e^{zt} dz = \frac{1}{2\pi i} \int_{s+} F(z) e^{zt} dz. \tag{12}$$

$F(z)$ and $f_0(t)$ are to be made subject to the same restrictions as $w(z)$ and $G(t)$ respectively.

Derivation of a Series for the Total Current

Let the amplifier be linear and of infinite power-carrying capacity. Let the output be connected to the input in such a way that the amplification ratio for one round trip is equal to the complex quantity AJ, where A is a function of the gain only and J is a function of ω only, being defined for all values of frequency from 0 to ∞.

---

[3] W. F. Osgood, "Lehrbuch der Funktionentheorie," 5th ed., Kap. 7, § 1, Hauptsatz. For definition of "analytic" see Kap. 6, § 5.

Let the disturbing wave $f_0(t)$ be applied anywhere in the circuit. We have

$$f_0(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(i\omega)e^{i\omega t}d\omega \tag{13}$$

or

$$f_0(t) = \frac{1}{2\pi i}\int_{s+} F(z)e^{zt}dz. \tag{13'}$$

The wave traverses the circuit and on completing the first trip it becomes

$$f_1(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} w(i\omega)F(i\omega)e^{i\omega t}d\omega \tag{14}$$

$$= \frac{1}{2\pi i}\int_{s+} w(z)F(z)e^{zt}dz. \tag{14'}$$

After traversing the circuit a second time it becomes

$$f_2(t) = \frac{1}{2\pi i}\int_{s+} Fw^2e^{zt}dz, \tag{15}$$

and after traversing the circuit n times

$$f_n(t) = \frac{1}{2\pi i}\int_{s+} Fw^ne^{zt}dz. \tag{16}$$

Adding the voltage of the original impulse and the first n round trips we have a total of

$$s_n(t) = \sum_{k=0}^{n} f_k(t) = \frac{1}{2\pi i}\int_{s+} F(1 + w + \cdots w^n)e^{zt}dz. \tag{17}$$

The total voltage at the point in question at the time t is given by the limiting value which (17) approaches as n is increased indefinitely [4]

$$s(t) = \sum_{k=0}^{\infty} f_k(t) = \lim_{n\to\infty} \frac{1}{2\pi i}\int_{s+} S_n(z)e^{zt}dz, \tag{18}$$

where

$$S_n = F + Fw + Fw^2 + \cdots Fw^n = \frac{F(1 - w^{n+1})}{1 - w}. \tag{19}$$

---

[4] Mr. Carson has called my attention to the fact that this series can also be derived from Theorem IX, p. 49, of his Electric Circuit Theory. Whereas the present derivation is analogous to the theory expressed in equations (*a*)–(*e*) above, the alternative derivation would be analogous to that in equations (*f*)–(*h*).

CONVERGENCE OF SERIES

We shall next prove that the limit $s(t)$ exists for all finite values of t. It may be stated as of incidental interest that the limit

$$\int_{s+} S_\infty(z)e^{izt}dz \tag{20}$$

does not necessarily exist although the limit $s(t)$ does. Choose M_0 and N such that

$$|f_0(\lambda)| \leq M_0. \qquad 0 \leq \lambda \leq t. \tag{21}$$

$$|G(t-\lambda)| \leq N. \quad 0 \leq \lambda \leq t. \tag{22}$$

We may write [5]

$$f_1(t) = \int_{-\infty}^{\infty} G(t-\lambda)f_0(\lambda)d\lambda. \tag{23}$$

$$|f_1(t)| \leq \int_0^t M_0Nd\lambda = M_0Nt. \tag{24}$$

$$f_2(t) = \int_{-\infty}^{\infty} G(t-\lambda)f_1(\lambda)d\lambda. \tag{25}$$

$$|f_2(t)| \leq \int_0^t M_0N^2tdt = M_0N^2t^2/2! \tag{26}$$

Similarly

$$|f_n(t)| \leq M_0N^nt^n/n! \tag{27}$$

$$|s_n(t)| \leq M_0(1 + Nt + \cdots N^nt^n/n!). \tag{28}$$

It is shown in almost any text [6] dealing with the convergence of series that the series in parentheses converges to e^{Nt} as n increases indefinitely. Consequently, $s_n(t)$ converges absolutely as n increases indefinitely.

RELATION BETWEEN $s(t)$ AND w

Next consider what happens to $s(t)$ as t increases. As t increases indefinitely $s(t)$ may converge to zero, indicating a condition of stability, or it may go beyond any value however large, indicating a runaway condition. The question which presents itself is: *Referring to (18) and (19), what properties of $w(z)$ and further what properties of $AJ(i\omega)$ determine whether $s(t)$ converges to zero or diverges as t increases*

---

[5] G. A. Campbell, "Fourier Integral," *B. S. T. J.*, Oct. 1928, Pair 202.

[6] E.g., Whittaker and Watson, "Modern Analysis," 2d ed., p. 531.

indefinitely? From (18) and (19)

$$s(t) = \lim_{n \to \infty} \frac{1}{2\pi i} \int_{s+} F\left(\frac{1}{1-w} - \frac{w^{n+1}}{1-w}\right) e^{zt} dz. \tag{29}$$

We may write

$$s(t) = \frac{1}{2\pi i} \int_{s+} [F/(1-w)] e^{zt} dz - \lim_{n \to \infty} \frac{1}{2\pi i} \int_{s+} [F w^{n+1}/(1-w)] e^{zt} dz \tag{30}$$

provided these functions exist. Let them be called $q_0(t)$ and $\lim_{n \to \infty} q_n(t)$ respectively. Then

$$q_n(t) = \int_{-\infty}^{\infty} q_0(t-\lambda)\phi(\lambda) d\lambda. \tag{31}$$

where

$$\phi(\lambda) = \frac{1}{2\pi i} \int_{s+} w^{n+1} e^{z\lambda} dz. \tag{32}$$

By the methods used under the discussion of convergence above it can then be shown that this expression exists and approaches zero as n increases indefinitely provided $q_0(t)$ exists and is equal to zero for $t < 0$. Equation (29) may therefore be written, subject to these conditions

$$s(t) = \frac{1}{2\pi i} \int_{s+} [F/(1-w)] e^{zt} dz. \tag{33}$$

In the first place the integral is zero for negative values of t because the integrand approaches zero faster than the path of integration increases. Moreover,

$$\int_I [F/(1-w)] e^{zt} dz \tag{34}$$

exists for all values of t and approaches zero for large values of t if $1 - w$ does not equal zero on the imaginary axis. Moreover, the integral

$$\int_C [F/(1-w)] e^{zt} dz \tag{35}$$

exists because

1. Since F and w are both analytic within the curve the integrand does not have any essential singularity there,
2. The poles, if any, lie within a finite distance of the origin because $w \to 0$ as $|z|$ increases, and
3. These two statements insure that the total number of poles is finite.

We shall next evaluate the integral for a very large value of t. It will suffice to take the C integral since the I integral approaches zero. Assume originally that $1 - w$ does not have a root on the imaginary axis and that $F(z)$ has the special value $w'(z)$. The integral may be written

$$\frac{1}{2\pi i}\int_C [w'/(1 - w)]e^{zt}dz. \tag{36}$$

Changing variables it becomes

$$\frac{1}{2\pi i}\int_D [1/(1 - w)]e^{zt}dw, \tag{37}$$

where z is a function of w and D is the curve in the w plane which corresponds to the curve C in the z plane. More specifically the imaginary axis becomes the locus $x = 0$ and the semicircle becomes a small curve which spirals around the origin. See Fig. 2. The function z and, therefore, the integrand is, in general, multivalued and the curve of integration must be considered as carried out over the appropriate Riemann surface.[7]

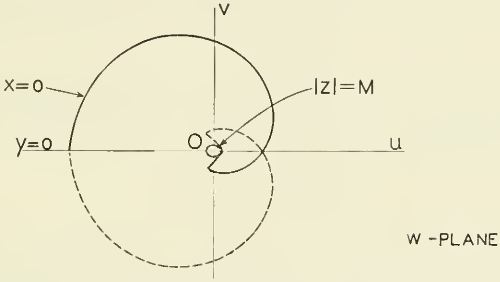


Fig. 2—Representative paths of integration in the w-plane corresponding to paths in Fig. 1.

Now let the path of integration shrink, taking care that it does not shrink across the pole at $w = 1$ and initially that it does not shrink across such branch points as interfere with its passage, if any. This shrinking does not alter the integral [8] because the integrand is analytic at all other points. At branch points which interfere with the passage of the path the branches stopped may be severed, transposed and connected in such a way that the shrinking may be continued past the branch point. This can be done without altering the value of the integral. Thus the curve can be shrunk until it becomes one or more very small circles surrounding the pole. The value of the total integral

---

[7] Osgood, loc. cit., Kap. 8.

[8] Osgood, loc. cit., Kap. 7, § 3, Satz 1.

(for very large values of t) is by the method of residues [9]

$$\sum_{j=1}^{n} r_j e^{z_j t}, \tag{38}$$

where z_j $(j = 1, 2 \cdots n)$ is a root of $1 - w = 0$ and r_j is its order. The real part of z_j is positive because the curve in Fig. 1 encloses points with $x > 0$ only. The system is therefore stable or unstable according to whether

$$\sum_{j=1}^{n} r_j$$

is equal to zero or not. But the latter expression is seen from the procedure just gone through to equal the number of times that the locus $x = 0$ encircles the point $w = 1$.

If F does not equal w' the calculation is somewhat longer but not essentially different. The integral then equals

$$\sum_{j=1}^{n} \frac{F(z_j)}{w(z_j)} e^{z_j t} \tag{39}$$

if all the roots of $1 - w = 0$ are distinct. If the roots are not distinct the expression becomes

$$\sum_{j=1}^{n} \sum_{k=1}^{r_j} A_{jk} t^{k-1} e^{z_j t}, \tag{40}$$

where A_{jr_j}, at least, is finite and different from zero for general values of F. It appears then that unless F is specially chosen the result is essentially the same as for $F = w'$. The circuit is stable if the point lies wholly outside the locus $x = 0$. It is unstable if the point is within the curve. It can also be shown that if the point is on the curve conditions are unstable. We may now enunciate the following

Rule: Plot plus and minus the imaginary part of $AJ(i\omega)$ against the real part for all frequencies from 0 to ∞. If the point $1 + i0$ lies completely outside this curve the system is stable; if not it is unstable.

In case of doubt as to whether a point is inside or outside the curve the following criterion may be used: Draw a line from the point $(u = 1, v = 0)$ to the point $z = -i\infty$. Keep one end of the line fixed at $(u = 1, v = 0)$ and let the other end describe the curve from $z = -i\infty$ to $z = i\infty$, these two points being the same in the w plane. If the net angle through which the line turns is zero the point $(u = 1, v = 0)$ is on the outside, otherwise it is on the inside.

If AJ be written $|AJ|(\cos\theta + i\sin\theta)$ and if the angle always changes in the same direction with increasing ω, where ω is real, the rule can be stated as follows: The system is stable or unstable according to whether or not a real frequency exists for which the feed-back ratio is real and equal to or greater than unity.

---

[9] Osgood, loc. cit., Kap. 7, § 11, Satz 1.

In case $d\theta/d\omega$ changes sign we may have the case illustrated in Figs. 3 and 4. In these cases there are frequencies for which w is real and greater than 1. On the other hand, the point (1, 0) is outside of the locus $x = 0$ and, therefore, according to the rule there is a stable condition.

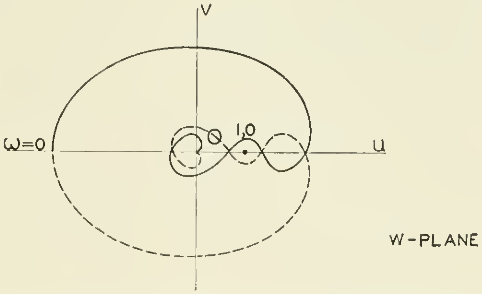


Fig. 3—Illustrating case where amplifying ratio is real and greater than unity for two frequencies, but where nevertheless the path of integration does not include the point 1, 0.

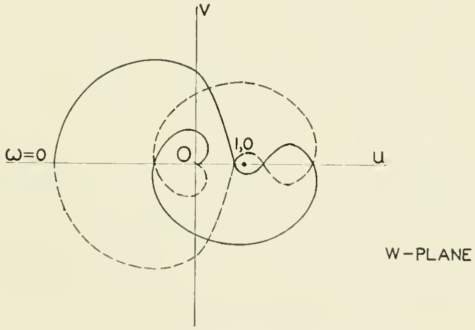


Fig. 4—Illustrating case where amplifying ratio is real and greater than unity for two frequencies, but where nevertheless the path of integration does not include the point 1, 0.

If networks of this type were used we should have the following interesting sequence of events: For low values of A the system is in a stable condition. Then as the gain is increased gradually, the system becomes unstable. Then as the gain is increased gradually still further, the system again becomes stable. As the gain is still further increased the system may again become unstable.

Examples

The following examples are intended to give a more detailed picture of certain rather simple special cases. They serve to illustrate the previous discussion. In all the cases F is taken equal to AJ so that f_0 is equal to AG. This simplifies the discussion but does not detract from the illustrative value.

1. Let the network be pure resistance except for the distortionless amplifier and a single bridged condenser, and let the amplifier be such that there is no reversal. We have

$$AJ(i\omega) = \frac{B}{\alpha + i\omega}, \tag{41}$$

where A and α are real positive constants. In (18) [10]

$$\begin{aligned} f_n &= \frac{1}{2\pi i} \int_I A^{n+1} J^{n+1}(i\omega) e^{i\omega t} d i\omega \\ &= B e^{-\alpha t}(B^n t^n / n!). \end{aligned} \tag{42}$$

$$s(t) = B e^{-\alpha t}(1 + Bt + B^2 t^2/2! + \cdots). \tag{43}$$

The successive terms f_0, f_1, etc., represent the impressed wave and the successive round trips. The whole series is the total current.

It is suggested that the reader should sketch the first few terms graphically for $B = \alpha$, and sketch the admittance diagrams for $B < \alpha$, and $B > \alpha$.

The expression in parentheses equals e^{Bt} and

$$s(t) = B e^{(B-\alpha) t}. \tag{44}$$

This expression will be seen to converge to 0 as t increases or fail to do so according to whether $B < \alpha$ or $B \geqq \alpha$. This will be found to check the rule as applied to the admittance diagram.

2. Let the network be as in 1 except that the amplifier is so arranged that there is a reversal. Then

$$AJ(i\omega) = \frac{-B}{\alpha + i\omega}. \tag{45}$$

$$f_n = (-1)^{n+1} B e^{-\alpha t}(B^n t^n / n!). \tag{46}$$

The solution is the same as in 1 except that every other term in the series has its sign reversed:

$$\begin{aligned} s(t) &= -B e^{-\alpha t}(1 - Bt + B^2 t^2/2! + \cdots) \\ &= -B e^{(-\alpha - B) t}. \end{aligned} \tag{47}$$

---

[10] Campbell, loc. cit. Pair 105.

This converges to 0 as t increases regardless of how great B may be taken. If the admittance diagram is drawn this is again found to check the rule.

3. Let the network be as in 1 except that there are two separated condensers bridged across resistance circuits. Then

$$AJ(i\omega) = \frac{B^2}{(\alpha + i\omega)^2}. \tag{48}$$

The solution for $s(t)$ is obtained most simply by taking every other term in the series obtained in 1.

$$\begin{aligned} s(t) &= Be^{-\alpha t}(Bt + B^3t^3/3! + \cdots) \\ &= Be^{-\alpha t} \sinh Bt. \end{aligned} \tag{49}$$

4. Let the network be as in 3 except that there is a reversal. Then

$$AJ(i\omega) = \frac{-B^2}{(\alpha + i\omega)^2}. \tag{50}$$

The solution is obtained most directly by reversing the sign of every other term in the series obtained in 3.

$$\begin{aligned} s(t) &= -Be^{-\alpha t}(Bt - B^3t^3/3! + \cdots) \\ &= -Be^{-\alpha t} \sin Bt. \end{aligned} \tag{51}$$

This is a most instructive example. An approximate diagram has been made in Fig. 5, which shows that as the gain is increased the

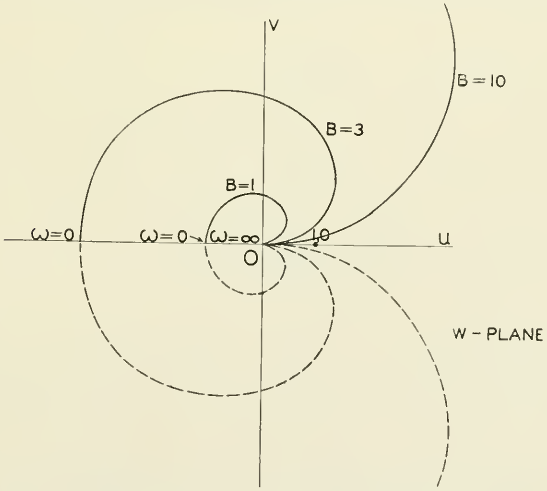


Fig. 5—Illustrating Example 4, with three values for B.

feed-back ratio may be made arbitrarily great and the angle arbitrarily small without the condition being unstable. This agrees with the expression just obtained, which shows that the only effect of increasing the gain is to increase the frequency of the resulting transient.

5. Let the conditions be as in 1 and 3 except for the fact that four separated condensers are used. Then

$$AJ(i\omega) = \frac{B^4}{(\alpha + i\omega)^4} \cdot \tag{52}$$

The solution is most readily obtained by selecting every fourth term in the series obtained in 1.

$$\begin{aligned} s(t) &= Be^{-\alpha t}(B^3t^3/3! + B^7t^7/7! + \cdots) \\ &= \tfrac{1}{2}Be^{-\alpha t}\,(\sinh Bt - \sin Bt). \end{aligned} \tag{53}$$

This indicates a condition of instability when $B \geqq \alpha$, agreeing with the result deducible from the admittance diagram.

6. Let the conditions be as in 5 except that there is a reversal. Then

$$Y = \frac{-\,B^4}{(\alpha + i\omega)^4} \cdot \tag{54}$$

The solution is most readily obtained by changing the sign of every other term in the series obtained in 5.

$$s(t) = Be^{-\alpha t}(-\,B^3t^3/3! + B^7t^7/7! - \cdots). \tag{55}$$

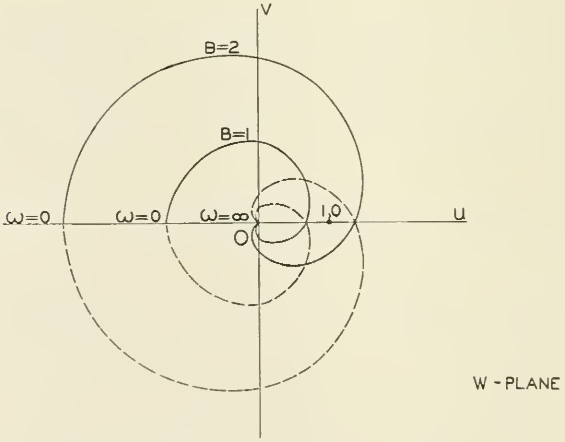


Fig. 6—Illustrating Example 6, with two values for B.

For large values of t this approaches

$$s(t) = -\tfrac{1}{2}Be^{(B/\sqrt{2}-\alpha)t}\sin(Bt/\sqrt{2} - \pi/4). \tag{56}$$

This example is interesting because it shows a case of instability although there is a reversal. Fig. 6 shows the admittance diagram for $B\sqrt{2} - \alpha < 0$ and for $B\sqrt{2} - \alpha > 0$.

7. Let

$$AG(t) = f_0(t) = A(1 - t), \qquad 0 \leq t \leq 1. \tag{57}$$

$$AG(t) = f_0(t) = 0, \qquad -\infty < t < 0, \qquad 1 < t < \infty. \tag{57'}$$

We have

$$\begin{aligned} AJ(i\omega) &= A\int_0^1 (1 - t)e^{-i\omega t}dt \\ &= A\left(\frac{1 - e^{-i\omega}}{\omega^2} + \frac{1}{i\omega}\right). \end{aligned} \tag{58}$$

Fig. 7 is a plot of this case for $A = 1$.

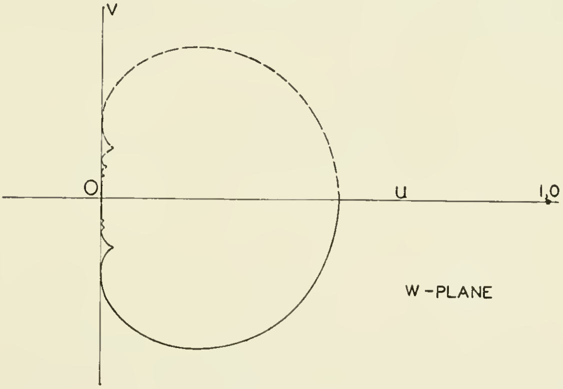


Fig. 7—Illustrating Example 7.

8. Let

$$AJ(i\omega) = \frac{A(1 + i\omega)}{(1 + i2\omega)}. \tag{59}$$

This is plotted on Fig. 8 for $A = 3$. It will be seen that the point 1 lies outside of the locus and for that reason we should expect that the system would be stable. We should expect from inspecting the diagram that the system would be stable for $A < 1$ and $A > 2$ and that it would be unstable for $1 \leq A \leq 2$. We have overlooked one fact, however; the expression for $AJ(i\omega)$ does not approach zero as ω

increases indefinitely. Therefore, it does not come within restriction (BI) and consequently the reasoning leading up to the rule does not apply.

The admittance in question can be made up by bridging a capacity in series with a resistance across a resistance line. This admittance obviously does not approach zero as the frequency increases. In any actual network there would, however, be a small amount of distributed capacity which, as the frequency is increased indefinitely, would cause the transmission through the network to approach zero. This is shown graphically in Fig. 9. The effect of the distributed capacity is essentially to cut a corridor from the circle in Fig. 8 to the origin, which insures that the point lies inside the locus.

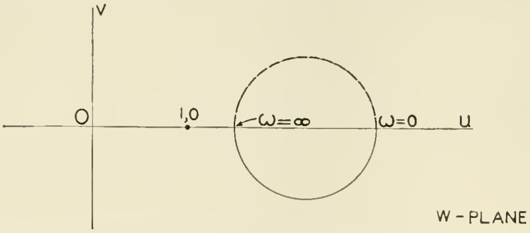


Fig. 8—Illustrating Example 8, without distributed constants.

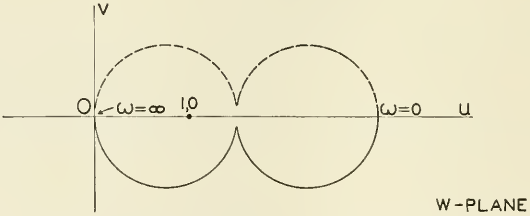


Fig. 9—Illustrating Example 8, with distributed constants.

Appendix I

Alternative Procedure

In some cases $AJ(i\omega)$ may be given as an analytic expression in $(i\omega)$. In that case the analytic expression may be used to define w for all values of z for which it exists. If the value for $AJ(i\omega)$ satisfies all the restrictions the value thus defined equals the w defined above for $0 \leq x < \infty$ only. For $-\infty < x < 0$ it equals the analytic continuation of the function w defined above. If there are no essential singularities anywhere including at ∞, the integral in (33) may be evaluated by the theory of residues by completing the path of integration so that all the poles of the integrand are included. We then have

$$s(t) = \sum_{j=1}^{j=n} \sum_{k=1}^{r_j} A_{jk} t^{k-1} e^{z_j t}. \tag{60}$$

If the network is made up of a finite number of lumped constants there is no essential singularity and the preceding expression converges because it has only a finite number of terms. In other cases there is an infinite number of terms, but the expression may still be expected to converge, at least, in the usual case. Then the system is stable if all the roots of $1 - w = 0$ have $x < 0$. If some of the roots have $x \geqq 0$ the system is unstable.

The calculation then divides into three parts:

1. The recognition that the impedance function is $1 - w$.[11]
2. The determination of whether the impedance function has zeros for which $x \geqq 0$.[12]

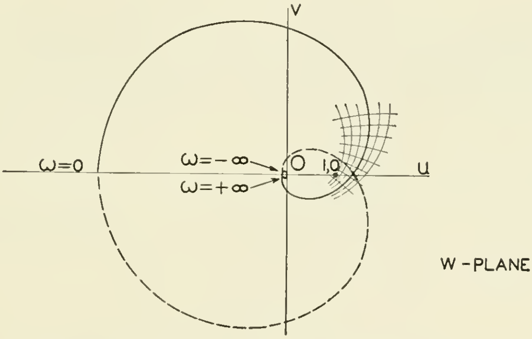


Fig. 10—Network of loci x = const., and y = const.

3. A deduction of a rule for determining whether there are roots for which $x \geqq 0$. The actual solution of the equation is usually too laborious.

To proceed with the third step, plot the locus $x = 0$ in the w plane, i.e., plot the imaginary part of w against the real part for all the values of y, $-\infty < y < \infty$. See Fig. 10. Other loci representing

$$x = \text{const.} \tag{61}$$

and

$$y = \text{const.} \tag{62}$$

---

[11] Cf. H. W. Nichols, *Phys. Rev.*, vol. 10, pp. 171–193, 1917.
[12] Cf. Thompson and Tait, "Natural Philosophy," vol. I, § 344.

may be considered and are indicated by the network shown in the figure in fine lines. On one side of the curve x is positive and on the other it is negative. Consider the equation

$$w(z) - 1 = 0$$

and what happens to it as A increases from a very small to a very large value. At first the locus $x = 0$ lies wholly to the left of the point. For this case the roots must have $x < 0$. As A increases there may come a time when the curve or successive convolutions of it will sweep over the point $w = 1$. For every such crossing at least one of the roots changes the sign of its x. We conclude that if the point $w = 1$ lies inside the curve the system is unstable. It is now possible to enunciate the rule as given in the main part of the paper but there deduced with what appears to be a more general method.

Appendix II

Discussion of Restrictions

The purpose of this appendix is to discuss more fully the restrictions which are placed on the functions defining the network. A full discussion in the main text would have interrupted the main argument too much.

Define an additional function

$$n(z) = \frac{1}{2\pi i}\int_I \frac{AJ(i\lambda)}{i\lambda - z}\,d(i\lambda), \qquad -\infty < x < 0. \tag{63}$$

$$n(iy) = \lim_{x \to 0} n(z).$$

This definition is similar to that for $w(z)$ given previously. It is shown in the theorem [13] referred to that these functions are analytic for $x \neq 0$ if $AJ(i\omega)$ is continuous. We have not proved, as yet, that the restrictions placed on $G(t)$ necessarily imply that $J(i\omega)$ is continuous. For the time being we shall assume that $J(i\omega)$ may have finite discontinuities. The theorem need not be restricted to the case where $J(i\omega)$ is continuous. From an examination of the second proof it will be seen to be sufficient that $\int_I J(i\omega)d(i\omega)$ exist. Moreover, that proof can be slightly modified to include all cases where conditions (AI)–(AIII) are satisfied.

---

[13] Osgood, loc. cit.

For, from the equation at top of page 298 [13]

$$\left|\frac{w(z_0-\Delta z)-w(z_0)}{\Delta z}-\frac{1}{2\pi i}\int_I\frac{AJ(i\lambda)}{(i\lambda-z_0)^2}d(i\lambda)\right|$$
$$\leq|\Delta z|\left|\frac{1}{2\pi i}\int_I\frac{AJ(i\lambda)d(i\lambda)}{(i\lambda-z_0-\Delta z)(i\lambda-z_0)^2}\right|,\qquad x_0>0.\qquad(64)$$

It is required to show that the integral exists. Now

$$\int_I\frac{AJ(i\lambda)d(i\lambda)}{(i\lambda-z_0-\Delta z)(i\lambda-z_0)^2}$$
$$=\int_I\frac{AJ(i\lambda)d(i\lambda)}{(i\lambda-z_0)^3}\left(1+\frac{\Delta z}{i\lambda-z_0}+\frac{\Delta z^2}{i\lambda-z_0}+\text{etc.}\right)\qquad(65)$$

if Δz is taken small enough so the series converges. It will be sufficient to confine attention to the first term. Divide the path of integration into three parts,

$$-\infty<\lambda<-|z_0|-1,\qquad -|z_0|-1<\lambda<|z_0|+1,\qquad |z_0|+1<\lambda<\infty.$$

In the middle part the integral exists because both the integrand and the range of integration are finite. In the other ranges the integral exists if the integrand falls off sufficiently rapidly with increasing λ. It is sufficient for this purpose that condition (BI) be satisfied. The same proof applies to $n(z)$.

Next, consider $\lim\limits_{x\to 0} w(z)=w(iy)$. If iy is a point where $J(iy)$ is continuous, a straightforward calculation yields

$$w(iy)=AJ(iy)/2+P(iy).\qquad(66a)$$

Likewise,

$$n(iy)=-AJ(iy)/2+P(iy)\qquad(66b)$$

where $P(iy)$ is the principal value [14] of the integral

$$\frac{1}{2\pi i}\int_I\frac{AJ(i\lambda)}{i\lambda-iy}d(i\lambda).$$

Subtracting

$$w(iy)-n(iy)=AJ(iy)\qquad(67)$$

If (iy) is a point of discontinuity of $J(iy)$

$$|w| \text{ and } |n| \text{ increase indefinitely as } x\to 0.\qquad(68)$$

Next, evaluate the integral

$$\frac{1}{2\pi i}\int_{x+I}w(z)e^{zt}dz,$$

---

[14] E. W. Hobson, "Functions of a Real Variable," vol. I, 3d edition, § 352.

where the path of integration is from $x - i\infty$ to $x + i\infty$ along the line $x =$ const. On account of the analytic nature of the integrand this integral is independent of x (for $x > 0$). It may be written then

$$\lim_{x \to 0} \frac{1}{2\pi i} \int_{x+I} w(z) e^{zt} dz = \lim_{x \to 0} \frac{1}{2\pi i} \int_{x+I} \frac{1}{2\pi i} \int_I \frac{AJ(i\lambda)}{i\lambda - z} e^{zt} d(i\lambda) dz$$

$$= \lim_{x \to 0} \frac{1}{2\pi i} \int_{x+I} \frac{1}{2\pi i} \lim_{M \to \infty} \left[\int_{-iM}^{iy-i\delta} + \int_{iy-i\delta}^{iy+i\delta} + \int_{iy+i\delta}^{iM} \right] \frac{AJ(i\lambda)}{i\lambda - z} e^{zt} d(i\lambda) dz$$

$$= \lim_{x \to 0} \left[\frac{1}{2\pi i} \int_{x+I} \frac{1}{2\pi i} \int_{iy-i\delta}^{iy+i\delta} \frac{AJ(i\lambda)}{i\lambda - z} e^{zt} d(i\lambda) dz + Q(t,\, \delta) \right], \; x > 0, \qquad (69)$$

where δ is real and positive. The function Q defined by this equation exists for all values of t and for all values of δ. Similarly,

$$\lim_{x \to 0} \frac{1}{2\pi i} \int_{x+I} n(z) e^{zt} dz$$

$$= \left[\lim_{x \to 0} \frac{1}{2\pi i} \int_{x+I} \frac{1}{2\pi i} \int_{iy+i\delta}^{iy+i\delta} \frac{AJ(i\lambda)}{i\lambda - z} e^{zt} d(i\lambda) dz + Q(t,\, \delta) \right], \; x < 0, \qquad (70)$$

Subtracting and dropping the limit designations

$$\frac{1}{2\pi i} \int_{x+I} w(z) e^{zt} dz - \frac{1}{2\pi i} \int_{x+I} n(z) e^{zt} dz = \frac{1}{2\pi i} \int_I AJ(i\lambda) e^{i\lambda t} d(i\lambda). \qquad (71)$$

The first integral is zero for $t < 0$ as can be seen by taking x sufficiently large. Likewise, the second is equal to zero for $t > 0$. Therefore,

$$\frac{1}{2\pi i} \int_{x+I} w(z) e^{zt} dz = \frac{1}{2\pi i} \int_I AJ(i\omega) e^{i\omega t} d(i\omega) = AG(t), \quad 0 < t < \infty \qquad (72)$$

$$-\frac{1}{2\pi i} \int_{x+I} n(z) e^{zt} dz$$

$$= \frac{1}{2\pi i} \int_I AJ(i\omega) e^{i\omega t} d(i\omega) = AG(t) \; - \infty < t < 0. \qquad (73)$$

We may now conclude that

$$\int_I n(iy) e^{iyt} d(iy) = 0, \qquad -\infty < t < \infty \qquad (74)$$

provided

$$G(t) = 0, \qquad -\infty < t < 0. \qquad \text{(AII)}$$

But (74) is equivalent to

$$n(z) = 0, \qquad (74')$$

which taken with (67) gives

$$w(iy) = AJ(iy). \qquad \text{(BIII)}$$

(BIII) is, therefore, a necessary consequence of (AII). (74′) taken with (68) shows that

$$J(iy) \text{ is continuous.} \qquad \text{(BII)}$$

It may be shown[15] that (BI) is a consequence of (AI). Consequently all the B conditions are deducible from the A conditions.

Conversely, it may be inquired whether the A conditions are deducible from the B conditions. This is of interest if $AJ(i\omega)$ is given and is known to satisfy the B conditions, whereas nothing is known about G.

Condition AII is a consequence of BIII as may be seen from (67) and (74). On the other hand AI and AIII cannot be inferred from the B conditions. It can be shown by examining (5), however, that if the slightly more severe condition

$$\lim_{y\to\infty} y^{\gamma}J(iy) \text{ exists,} \qquad (\gamma > 1), \qquad \text{(BI}a\text{)}$$

is satisfied then

$$G(t) \text{ exists,} \qquad -\infty < t < \infty, \qquad \text{(AI}a\text{)}$$

which, together with AII, insures the validity of the reasoning.

It remains to show that the measured value of $J(i\omega)$ is equal to that defined by (6). The measurement consists essentially in applying a sinusoidal wave and determining the response after a long period. Let the impressed wave be

$$E = \text{real part of } e^{i\omega t}, \qquad t \geqq 0. \qquad (75)$$

$$E = 0, \qquad t < 0. \qquad (75')$$

The response is

$$\text{real part of} \int_0^t AG(\lambda)e^{i\omega(t-\lambda)}d\lambda$$

$$= \text{real part of } Ae^{i\omega t}\int_0^t G(\lambda)e^{-i\omega\lambda}d\lambda. \qquad (76)$$

For large values of t this approaches

$$\text{real part of } Ae^{i\omega t}J(i\omega). \qquad (77)$$

Consequently, the measurements yield the value $AJ(i\omega)$.

[15] See Hobson, loc. cit., vol. II, 2d edition, § 335. It will be apparent that K depends on the total variation but is independent of the limits of integration.

Contemporary Advances in Physics, XXIII

Data and Nature of Cosmic Rays

By KARL K. DARROW

"Cosmic rays" is the name of the ultimate cause which maintains that part of the ionization of the air which cannot be ascribed to the rays of radioactive substances on earth nor to any other known agency. The measurement of this residue, the discrimination between it and that part of the ionization which is due to familiar rays, is the first problem of cosmic-ray research. Second comes the problem of learning, from measurements made at as many places and under as many conditions as possible, the nature of the mysterious ionizing agent. One naturally begins by assuming it to be like in kind to one or another of the known types of ionizing rays, but different in quality; *e.g.* to consist of electrons faster than any known electrons, or photons of greater energy and lesser wave-length than any known photons. It is not yet certain whether one of these hypotheses will fit, or rays of some new type must be imagined. The direct measurements of ionization are supplemented by observations of material particles of evidently enormous energy which dart across the atmosphere in straight paths leaving trains of ions behind. On the whole it seems highly probable that these particles, or those (if such there be) from which they receive their energy, come to the earth from outer space; and the energy which they bear is so great, that its source must be some process not yet known.

THE subject of this article is unique in modern physics for the minuteness of the phenomena, the delicacy of the observations, the adventurous excursions of the observers, the subtlety of the analysis, and the grandeur of the inferences. The effect which is studied, which may be described as the liberation of electrons from the molecules of the air by agents otherwise unknown, amounts at sea-level to the liberation of only about 1500 of these per cubic decimetre of air per second. It is not the whole of the observed effect; these 1500 electrons are those which are left over out of a quantity often much greater, after allowance is made for the actions of all known ionizing agents. The methods employed are ranked among the most ingenious and sensitive of science; yet the apparatus is not invariably set up in the calm seclusion of the laboratory. Physicists with their frail machines have gone to high mountain ponds in the Sierras and the Andes, to the distant wildernesses about the earth's magnetic poles; they have scooped out cavities in Alpine glaciers, they have lifted hundredweights of lead to the tops of peaks above the snow-line, they have cruised the arctic and the tropical oceans, they have descended into tunnels and deep mines, they have ascended into the sky in aeroplanes and balloons. As for the analysis of the precious data obtained with so much labor, it will be evident from this article how intricate a process it now proves to be.

And the incentive for all these labors? At first, the problem seems little more exciting than the tracing of an insignificant leak in an apparatus supposedly insulated. There is, however, reason to suspect that *this* leak is due to causes really sensational. The comprehended part of the ionization of the air being due to known rays such as X-rays, light, and the radiations from radioactive substances, it is natural to ascribe the mysterious residue to rays as yet unknown. It turns out then that these hypothetical rays must lie beyond any yet discovered; if they are electrons, they must be swifter electrons—if protons, faster protons—if corpuscles of light, then corpuscles of higher frequency and higher energy, than any thus far known; or else they must be particles of a totally new variety. It is not easy to imagine where they could come from on earth, and there are various reasons for supposing that they wander in from outer space. Messengers from the depths of space and from the stars—corpuscles of visible light, meteorites, the electrons which are the presumptive causes of the aurora—are constantly being received and have been amply interpreted; it is probable that these cosmic rays have a message of value, perhaps of the first importance.

The dates of the beginnings of cosmic-ray research have been much disputed, in an unprofitable fashion. I like to set aside the controversies by choosing a time extravagantly remote, the year 1785! In that year, it is said, Coulomb made the first acceptable proof that the air of the atmosphere is conductive. It had long been known that an electrified body, mounted on the best available insulator and set up in the air, slowly loses its charge; it had, however, been thought that the "leak" is partly an escape through the insulator or over the surface thereof, partly a carrying-off of charge by particles of dust which drift up to the electrified body, touch it and depart—Coulomb found that in addition to these, there is an actual conduction through the air.

After this discovery, progress ceased for many years. Towards the end of the nineteenth century it was, however, learned that the air becomes much more conductive than normally it is, when rays of certain kinds pervade it—X-rays for instance, and the radiations from radioactive atoms. Further, it was found that some part, at least, of the normal conductivity of atmospheric air is due to rays from radioactive atoms blown about by the wind or embedded in the ground. Then came the time for asking: Is there a residual part of the normal conductivity of the air, which cannot be ascribed to rays of any kind hitherto known? With the asking of this question, the way was opened to the sequence of researches, which form the topic of this article. "Cosmic rays" are *by definition* the cause of that residue of the ionization of the atmosphere, which is left over after deduction of all of the ionization which can be traced to the action of the rays of radioactive substances, or any other known cause whatever.

I write "ionization" instead of "conductivity"; this is because it is well enough known that when a volume of gas is and remains conductive, it is constantly being ionized—which is to say, electrons are constantly being detached from its molecules. Molecule and electron become ions, the members of an "ion-pair"; if there is an electrified body nearby, those of the sign opposite to its charge are drawn to it (those of the other sign going to the earth) and step by step their charges neutralize its own. The question set above may therefore be rephrased: Are ion-pairs appearing in the atmosphere oftener than rays of known varieties are making them?

To answer this question, one must of course exclude all of the rays from radioactive substances (and, naturally, X-rays and other artificial kinds) from the volume of air which one is studying; or if they cannot be excluded altogether, one must estimate the ionizing-power of those which remain. They are important in cosmic-ray research as causes of error; they are important also in another way, as tentative models for theories of the cause of the residual ionization. In attempting to explain this mysterious residue, the procedure both natural and wise is to begin by supposing it similar to alpha-rays, or else to beta-rays, or else to gamma-rays. In the end it may be found needful to postulate some ionizing-agent entirely different from all of these; but in the beginning, they should be the guides.

Of the three named classes of rays, one (the alpha-particles or positive corpuscles) has a sharply-limited range in air so short, that such as come from the ground are stopped within a few inches; and a range in metal so extremely short, that even if one wished it would not be possible to build a durable air-chamber with solid walls so thin that the alpha-rays could get through.

The second (the beta-particles or fast electrons) have a range which varies from one radioactive element to another, and even from one particle to another among those emitted by a single element. This range is often much longer than that of an alpha-particle, and less precise. Air at a distance of many metres above the ground may be sensibly ionized by beta-corpuscles proceeding from the soil. All along their paths, beta-particles engender ion-pairs profusely—in air at normal (sea-level) density, never fewer than forty per cm. Nevertheless they do not enter an air-chamber walled by a very few centimeters of lead, and need not be a serious cause of confusion.

More formidable is the third kind, the gamma-rays or short electromagnetic waves. They also vary from one element to another, and different wave-lengths may proceed from a single element; some are innocuous, but some are much more penetrating than even beta-rays. One must be careful about the language here; one must not say that they have a great range. To speak of "range" implies that when a number of particles is started off side by side with equal speed along parallel paths into a stratum of matter, they continue to follow parallel paths—albeit with diminishing speed—until they are brought to a stop, this termination coming for each of them at the same distance from their source. Of alpha-particles this is nearly a true statement, of beta-particles it is a fair approximation. Of gamma-corpuscles or photons (for gamma-rays are high-frequency light, and may therefore be regarded alternatively as corpuscles and as waves) it is not true at all. If a number of these, all of identical wave-length, is started off side by side along parallel paths into a stratum of matter, one after another impinges on an atom, or (it is probably better to say) upon an electron belonging to some atom. Until such an impact, the photon does not ionize at all, but passes unperceived. The number which at a distance x from the source have not yet suffered an impact varies as an exponential function $e^{-\mu x}$ of that distance.

At such an impact the gamma-ray photon loses energy and suffers a deflection, which may or may not be considerable. It also liberates the electron, which is an act of ionization; but usually the electron itself, being started off with considerable speed, liberates a great number of additional electrons from other atoms of the gas, so that the direct ionization by the gamma-rays falls far behind the indirect; *most of the ionization by gamma-rays is done by the intermediacy of fast electrons.* One may say that the lower the value of μ, the more penetrating, or "harder," are the gamma-rays; one may compare values of μ for gamma-rays of different wave-lengths; but one may not compare a value of μ for these with a value of range for beta-particles or alpha-particles (unless one should arbitrarily define the range of a beam of gamma-rays as the distance over which its intensity is reduced by say 99 per cent, which would be very misleading).

These are very important matters, for the chance of deciding whether the cosmic rays are likest to gamma-rays or to material particles turns largely upon differences such as these. Another detail is important: the practice of stating values of μ for standard materials, such as water or lead. It is *approximately* true that slabs of various materials will absorb the same fraction of an incident beam of gamma-rays (of a single wave-length, be it understood) if their thicknesses are so adjusted that all of them have the same mass per unit of surface-area. The densities of air at sea-level, of water and of lead stand to each other in about the ratios .0013 : 1 : 11.4; a metre of water is therefore "equivalent" in this respect to about nine centimetres of lead or three-quarters of a kilometre of air at sea-level. After measuring μ in a material of density ρ, one may reduce it to a value approximately equal to that which would be found in a standard material (lead, for instance) of density ρ_0, by multiplying it with the factor (ρ_0/ρ). A better approximation yet, for the harder gamma-rays, is the statement that the slabs absorb the same fraction of the gamma-ray beam if the atoms under each unit of surface-area have the same number of bound electrons altogether. Denoting the number of bound electrons per unit volume of the material in question by E, the corresponding number for lead (say) by E_0, we find (E_0/E) for the value of the factor by which an observed value of μ is to be multiplied, in order to convert it into one approximately valid for lead.[1] To give a definite example, I will mention simply that ten centimetres of lead intercept all but a few per cent of the hardest gamma-ray corpuscles which come from radioactive bodies; and that a couple of kilometres above the earth, one need not worry about the influence of the radioactivity of the ground.[2]

So, to avoid ionization of air by the rays of the known radioactive elements at the surface of the earth, one surrounds the air which one is observing by matter more than thick enough to stop all of the alpha and all of the beta-particles, and thick enough to cut off all but a few per cent of the gamma-corpuscles. Often this is done by using plates of lead several cm. thick for the walls of the air-chamber. Sometimes it is done by sinking the chamber into a lake, or digging a hole into the ice of a glacier. A much greater thickness of water frozen or liquid is required to intercept the unwanted rays, than of lead; this thickness must surround the air-chamber on all sides, towards the bottom of the lake, towards its edges, even (if there be overhanging mountains, or radioactivity in the air) towards its surface; but water is inexpensive, and does not have to be carried about. Or by going up in a balloon, or sending up the air-chamber in a balloon without oneself attending it, one may put enough air beneath the apparatus to screen away the rays from the ground.

But are there not radioactive atoms in the very materials used to protect the ionization-chamber? the air, the ice, the water, the leaden walls themselves? Of the atmospheric air I have already admitted this; the quantity of radioactive gas commingled with it can, however, be measured, is generally very small, and does not seem to bother observers when they are not too close to the ground. As for the water: Millikan has always carefully chosen "snow-fed" lakes, high in the mountains where the water is derived not from springs which have seeped through soil, but from snow which has fallen onto bare rock from on high. Ice of glaciers, it is to be presumed, has an origin equally uncorrupted. The same cannot be said for the water of Lake Constance wherein Regener's data were taken, but he by special tests proved the effect of its radioactivity to be slight.

---

[1] For an element of atomic weight A and atomic number Z and density ρ, the value of E is $Z\rho/Am_H$, the symbol m_H standing for the mass of the hydrogen atom. For lead $Z = 82$, $A = 206$, $\rho_0 = 11.4$. For a compound of known constitution the formula can easily be worked out; if the constitution is unknown but certainly involves atoms of high atomic number only, one may put $Z/A = \frac{1}{2}$ as a rough approximation.

[2] Millikan in a specific case says that 4.8 cm. of lead intercept 90 per cent of the gamma-rays from igneous rocks (*Phys. Rev.* (2), **28**, p. 862; 1926).

As for the lead (or whatever other metal or metals may be used in making the ionization-chamber) this too may be contaminated with radioactive atoms. Some people have mounted wire netting all around the interior of the chamber at some distance from the walls, so that such rays as come from the wall and are soft (in the sense opposite to that in which I spoke of gamma-rays as hard) may be largely absorbed between the wall and the net; voltages are so arranged that the ions produced in this space are not counted by the observer. This device is not always used, nor when used is it fully effective. A somewhat experienced man might expect to be able to allow for the radioactivity of the wall, through knowledge acquired by varying the density of the air and studying the ionization as function of density; this has been done, but we shall see that the results have added new mysteries to the problem.

The walls, then, must be expected to cause a permanent ionization in the air-chamber, constant and independent of the outside world. Moreover electrical leaks must occur, whereby the charged electrode gradually loses its charge by conduction through the not-quite-perfect insulators on which it is mounted. How can we hope to distinguish the joint effect of these from that of the cosmic rays? or, to put the same question more properly, how can we hope to find what part of the ionization, if any, is not due to either of these?

This is indeed a serious question, inasmuch as it is a problem not of distinguishing a known from an unknown, but two unknowns from one another. Of course, if one makes the gratuitous assumption that the "cosmic rays" cannot penetrate x feet of rock, then one may take the apparatus into a cave or hollow under more than x feet of rock, and measure the rate at which the charged electrode is discharged; and then, after deducting the allowance to be made for the rays from radioactive substances in the rock which penetrate into the chamber, one can say that the residue (*Restgang*, the Germans call it) is the joint effect of the causes aforesaid, and must thenceforth be subtracted from every reading made with the same apparatus. Kolhörster did this on a grand scale, going many hundreds of metres down into one of the potash-mines of Stassfurt, where there is a great hollow excavated in rocksalt; the thickness of the overlying rock was great, but as a drawback there was a powerful radioactivity of the potassium in the rock-salt, causing hard gamma-rays of which a perceptible fraction invaded the chamber. In Java the physicist Clay took his machinery into a tunnel, where he was covered by rock eighty-four metres thick and far less rich than rocksalt in radioactive atoms; in the Sierra Nevada Millikan descended 185 metres into granite.

More instructive is the customary procedure of Millikan and of Regener. We shall presently consider curves (Figs. 3, 4, 5) which show the rate of discharge of electroscopes in air-chambers sunk under water, plotted against the depth of submergence. The discharge-rate falls off as the air-chamber is lowered, but the decline grows slower and slower; the curve flattens out and seems to approach, seems even to attain, a certain horizontal line. It is the ordinate of this line which is taken as the *Restgang;* this is subtracted from all the other ordinates, the differences are ascribed to the cosmic rays. Now if one could be absolutely sure that this is the line to which the curve is making asymptotic approach, there would be no uncertainty. But since no measurement is perfectly precise, no one can say absolutely that the curve is not still gently sloping, towards an asymptote distinctly lower than the lowest value measured. Yet it is on some guess as to the answer of this unanswerable question that there rest, not (fortunately) the proof that there are such things as cosmic rays, but the estimates of their amount and of their greatest penetrating power. I revert to this question later; at present I merely mention it, in order to show how difficult these estimates may be.

A very important point now demands to be noticed. In taking these elaborate precautions to exclude the rays from radioactive substances beyond the walls of the air-chamber, is one not also keeping out some of the cosmic rays themselves? To deny this would be all but impossible; it would amount to assuming that *all* ionizing rays, apart from those which we recognize as proceeding from known radioactive atoms, are so much more penetrating than these that they pass absolutely undiminished through centimetres of lead and metres of water—an assumption which has only to be stated, to show itself improbable. True, *if* the unknown rays come altogether from above, and *if* the radioactivity of the air can be allowed for, one may omit to carry the screen of lead over the top of the air-chamber—or, one may float the apparatus on or just below the surface of the lake—or, one may rely exclusively on observations made in the upper air. But it is likely that part of the unknown radiation comes very obliquely downward through the atmosphere, and therefore is liable to be obstructed by lateral walls of lead which cannot be removed without admitting to the apparatus other rays which come very obliquely upward from the earth. So, every sort of screening by water or by metal probably keeps out some of the rays which are wanted, along with those which are unwanted. Only the observations in the uppermost air may give the full effect of the mysterious radiation, and perhaps not even they, for the walls of the chamber cannot be reduced to infinite thinness.

How great is the ionization left over to be ascribed to cosmic rays, after *Restgang* and radioactivity are allowed for? The absolute value is still a subject of controversy; fortunately it is a minor matter, knowledge of which is not required for solving the major problems; thus, to determine the penetrating power of the rays one needs only *relative* values at various depths of water. Concerning it I will merely state, that *in the atmosphere near sea-level, the number of ions of either sign appearing per cubic centimetre per second, and attributed to cosmic rays, averages about one and one-half.* Say we have an ionization-chamber of volume U (in cubic centimetres), at sea-level, filled with air at atmospheric pressure. Its charged electrodes will be discharged at a certain rate; part of this rate of discharge can be explained in the ways aforesaid; the rest, the inexplicable residue, will amount to about $1.5\,Ue$. This residue is the evidence for the cosmic rays. The electron-charge e is so small, that even when the volume of the chamber is thousands of cubic centimetres the product ($1.5\,Ue$) is no great quantity of electricity. High up, the effect is much greater; according to Piccard it is so great at sixteen kilometres over sea-level, that if the air up there were as dense as it is by the sea, there would be 200 ions of either sign appearing per cc. per second.

Stating the effect in ions per cc. per second has one disadvantage: it suggests that in each cubic centimetre of an ionization-chamber the ions are created at random,—one in the middle, perhaps, at a certain instant, the next somewhere else two-thirds of a second later, the next yet somewhere else two-thirds of a second later yet, and so on. What happens, however, is this: at practically a single instant hundreds, or even thousands, of ion-pairs are created along a single straight line traversing the chamber; this is followed by many seconds, or even minutes, during which nothing happens; then there is another such event, a train of ions suddenly appearing along a straight line, not however the same one; another interval, another train of ions, and so forth. The method of measurement which thus far I have been presupposing conceals these individual events, giving merely the sum of all the ions which come into being over a period of hours (or, at least, of many minutes). Towards the end of the article I shall speak of other, and very striking, methods which reveal them.

Passing now to the experiments, I will take up first the various groups of measurements made, each with a single instrument of the ionization-chamber or electroscope type, at various heights above the ground or various depths below the surfaces of lakes.

Imagine, to begin with, a metal-walled air-filled box, having inside it a pair of strips of gold-foil hanging side by side from a metal knob at the end of an insulating rod, and in one wall a window through which this "gold-leaf electroscope" can be seen. The metal wall is earthed, and the leaves of gold are charged, usually by a metal rod so mounted in a ground-joint in the wall that by turning the joint the rod can be touched to the knob which holds the leaves. The wall is disconnected from earth, the box is set up in whatever place is chosen for the measurement. The divergence of the gold-leaves is measured at the beginning and at the end of a chosen interval of time. By calibrating the device (I will not enter into the details of this process) one may determine how much charge of one sign has been lost from the leaves, how much charge of the opposite sign has been gained by the leaves, in the time allotted. This is the ultimate datum. The process is then repeated in other places.

All of the experiments which I shall now describe are in principle like this imagined one, but greatly improved in technical detail. Gold-leaves would be too frail for an apparatus meant to be moved about; for them are substituted sometimes a pair of quartz fibres brought together at both ends, not at one only (Fig. 1) or else a "string electrometer." Sometimes the observations are made fifteen kilometres up in the air, or two hundred metres down under water. It is not always convenient, sometimes not feasible, for the observer to go with his instrument; it is not always safe to make the initial reading before the apparatus is sent on its way, the final reading after it is brought back. The observations must then be delegated to a mechanism; the divergence of the fibres or the position of the string is recorded by photography, perhaps at intervals of time previously chosen and regulated by machine, perhaps continuously. The pieces of apparatus designed to do these things, and meanwhile to survive immersion or transportation to great heights and low temperatures, are often prodigies of compactness.

Though the work of Millikan and his school is unrivalled for extent, I will begin with Regener's, which is an unique example of a wide range of water-depths covered in a single experiment. The ionization-chamber (walls of steel a centimeter thick, 39 liters (!) in volume filled with carbon dioxide at a pressure of 30 atmospheres) was floated for days at a time at each of several levels in the great Lake of Constance (*Bodensee*) on the northern boundary of Switzerland. Every hour a light flashed for a few seconds, and the image of the fibre of the string-electrometer was impressed upon a film, wound about a slowly and equably revolving drum. One sees in Fig. 2 the successive images of the string, developed after the experiment was completed and the film removed from the drum; it is evident how they lie closer together as the chamber is lowered deeper and deeper beneath the surface, the ionization therefore becoming feebler and feebler. The

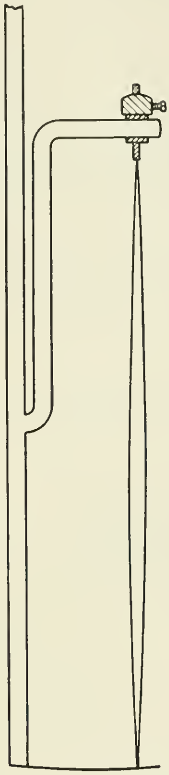


Fig. 1—Sketch of a two-fibre electroscope (after Millikan).

water, as we shall see, is serving not merely as a screen for keeping out the rays of radioactive matter in the earth but as an instrument for studying the penetrative power of the unknown rays themselves. Records, as I just said, were made an hour apart, and ordinarily were continued over two days at each depth.

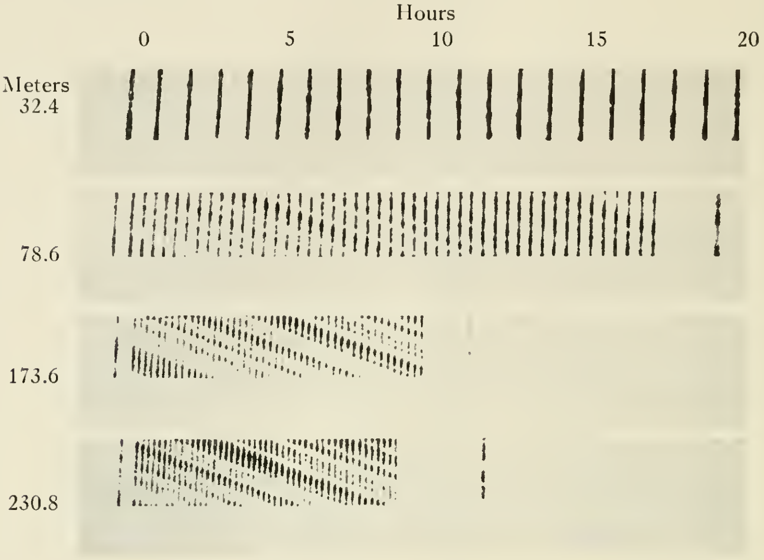


Fig. 2—Photographic registration of the position of the string of a string electrometer, showing gradual shift due to discharge by cosmic rays (Regener).

Measurements were made at seven depths: in metres, these amounted to 32.4, 78.6, 105.2, 153.5, 173.6, 186.3, and 230.8. They were plotted as function of depth; the resulting curve, *shifted downward by 0.78* (we need not consider the exact meaning of the units along the axis of ordinates) appears as *J* in Fig. 3. It is concave-upward, of a shape which suggests that there are both a *Restgang*, and ionizing-rays diminishing in strength as they descend through the water from above. But it is not sufficiently prolonged to permit the eye to decide with confidence the value of the *Restgang;* the exact effect of the ionizing rays is therefore not to be judged by inspection. How did Regener proceed? He postulated that the readings y at the three greatest depths (depth measured downward from the surface being denoted by x) are sums of a constant and an *exponential* term:

$$y = A + Be^{-\mu x}; \; x = 173.6, 186.3, 230.8,$$

and found what values must be given to A, B, and μ in order to get the best possible fit of such an expression to the data. He then noticed that the same expression would equally well fit the data for the next two of the depths, 153.5 and 105.2; and this convinced him that it is physically sound.

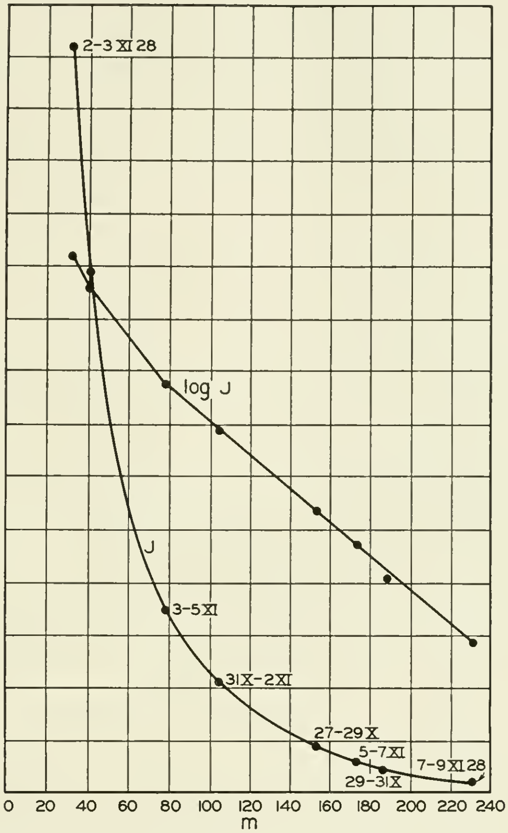


Fig. 3—Ionization ascribed to cosmic rays as function of depth beneath surface of Lake Constance (the straighter curve is a semi-logarithmic, the other a direct plot) (Regener).

The ionization has therefore been separated by assumption into a *Restgang* A and an effect of ionizing rays diminishing exponentially with increase of depth. The value of A is 0.78; it is the downward shift mentioned just above, so that the curve J in Fig. 3 shows the part of the ionization ascribed to cosmic rays. The other curve shows the logarithms of the ordinates of the curve J; the latter being exponential, the former is straight; it is the close fit of this line to the circles along its path (the logarithms of the data) that for the eye is the most striking evidence for Regener's procedure.

Now, of the three known kinds of rays from radioactive substances, one only is characterized by an exponential decline of its strength with increase of thickness of matter traversed: to wit, the gamma-rays. Regener was therefore assuming, as had many others before him, that *cosmic rays which have penetrated many metres of water behave like a beam of monochromatic gamma-radiation, corpuscles of light of a single frequency and energy, coming vertically from above.* The value of μ—.00018 in this case[3]—is many times smaller than that of any known variety of gamma-rays; this is expressed by saying that the cosmic rays are considerably more penetrating or "harder" than the hardest of gamma-rays.

While the expression $(A + Be^{-\mu x})$ fits tolerably the five experimental points at the five greatest depths, it misses the two others, giving for 32.4 and 78.6 metres lower values of ionization than those observed. One might add a second exponential term with a different, much larger value of μ; thus in effect assuming that the cosmic rays behave like a mixture of gamma-rays of two different wave-lengths, corpuscle-energies and hardnesses. This in fact is the custom, though Regener did not do it in the paper whence I am quoting, being interested mainly in the far end of the curve—in the range of depths at which, as is commonly said, *all but the hardest component of the cosmic rays have already been filtered out.* For, if a function of x is the sum of a number of terms of the form $B_i e^{-\mu_i x}$, with differing values of the constants μ_i: then as x is increased the various terms fall off, but in such a way that eventually the one with the lowest μ predominates more and more over all the rest, however small may be its coefficient B. This is "filtering," a process of great use in the science and practice of radioactivity; many of the radioactive elements, or of the mixtures thereof which are common, send out gamma-rays of various degrees of penetrating-power, among which the hardest may be almost isolated from the rest (at the price, of course, of a great cutting-down of its own intensity) by a thick screen of absorbing matter. Even if the screen is not thick enough to make the hardest component predominant, it may alter greatly the relative proportions of harder components and softer. The rays responsible for the 1.5 ions at sea-level, if of the nature of gamma-rays, must have undergone a good deal of this alteration and "hardening" in passing through the atmosphere; not nearly so much, however, as those deep down in Lake Constance, the whole of the atmosphere being no more powerful an absorber than a layer of ten metres of water.

We turn now to the great series of experiments made by Millikan and his school. Fig. 4 embodies the data of three, in each of which an air-chamber was lowered into a high-lying snow-fed[4] lake (Millikan's reasons for choosing such I have already stated). Crosses stand for data obtained in Arrowhead Lake, California (altitude 1550 metres); circles correspond to Muir Lake, California (3580 metres); the dots to which arrows point are the reward of a long hard journey, for they show data from a lake—Miguilla—no less than 4570 metres above the sea, in the Bolivian Andes. All of the measurements here plotted were taken at depths of a metre or more below the water surface, but none so great as Regener's—five metres was the maximum depth at Miguilla, fifteen at Arrowhead and twenty at Muir.[5]

---

[3] Expressed in cm.$^{-1}$; in (metres)$^{-1}$, Millikan's customary unit, it would be .018.

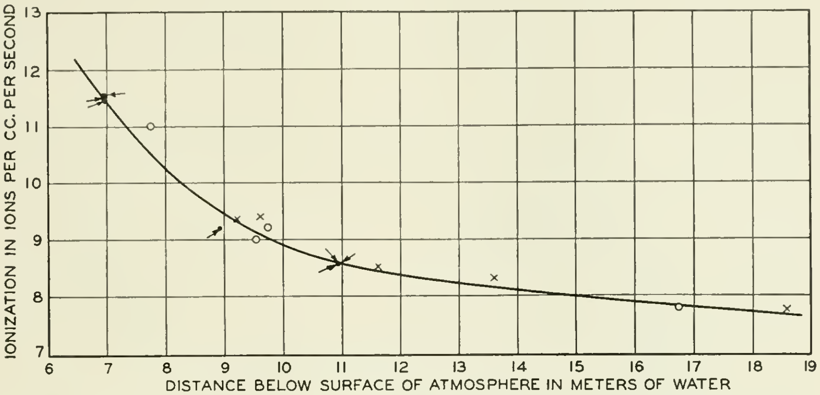


Fig. 4—Ionization ascribed to cosmic rays as function of depth beneath top of atmosphere (early curve of Millikan and Cameron).

The points of the three sets lie close to a single smooth curve. Each of the three, however, has been shifted horizontally, to make allowance for the different depths of air overlying the three lakes: it has been assumed that the air over Miguilla is the equivalent of 5.95 metres of water, over Muir to 6.75 and over Arrowhead to 8.6 metres. To figure out the equivalents it is necessary to know the distribution-in-height of the air (Millikan got it from the Smithsonian Institution tables) and the ratio of the absorbing-power of air of given density and the absorbing-power of water for the cosmic rays. This last not being known *a priori*, it is the custom to take the inverse ratio of the densities of the media (page 152)—another instance of assuming that cosmic rays are of the nature of gamma-rays. The assumption is strengthened by the success of the thus-made allowance in bringing all the points upon a single curve.

---

[4] Regener of course admits that Lake Constance is not of this character, but by later experiments (briefly mentioned in his note in the *Physikalische Zeitschrift*) he found that radioactivity of its water was not distorting his results.

[5] Data were obtained at lesser depths than a metre, but the values of ionization are higher than would be inferred by prolonging the curve which is valid for greater depths. This recalls a feature of Regener's data, but here the excess is ascribed by Millikan not to soft cosmic rays, but to the radio-activity of the overlying air and the nearby mountains. Of the points plotted in Fig. 4, the "two which fall farthest from the curve correspond to single readings, and hence should be given little weight in comparison with points which represent the means of three or four readings" (Millikan & Cameron).

The data of Fig. 4 were obtained in 1925 and 1926; those of the following Fig. 5, in 1927. New readings were made in the waters of Arrowhead Lake and in another not previously used, Gem Lake, 2,760 metres up in the high Sierras; into this last the ionization-chamber was lowered to a depth as great as sixty metres. The data for the two localities were altered to make allowance in the foregoing way for the different heights of air overlying the two lakes, and the values so obtained were plotted against "depth in metres-of-water beneath top of atmosphere." Evidently the data fit better to the curve than in the prior work: this Millikan and Cameron ascribe to improvements in the new ionization-chamber or "electroscope" (outwardly like that in Fig. 6) into which air was compressed to a pressure of *eight* atmospheres instead of *one*, as previously. Incidentally, in Fig. 5 the *Restgang* appears under the title "zero of the electroscope."

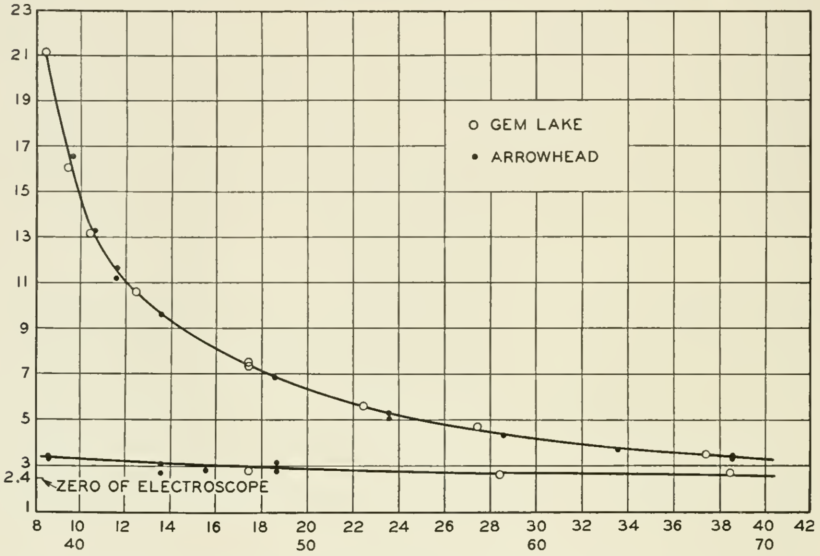


Fig. 5—Ionization ascribed to cosmic rays as function of depth beneath top of atmosphere (later curve of Millikan and Cameron).

The observations had thus been extended over the range, of metres-of-water beneath the "top of the atmosphere," from 8.45 to 69. During the three following years Millikan and Cameron made yet another and more-closely-spaced aggregate of readings, over the same range with slight extensions at both ends, in the same two lakes, with an electroscope (Fig. 6) into the 1622 cubic centimetres of whose volume air was forced to a pressure of 30 atmospheres. The reader can find the new curve in the issue of the *Physical Review* for February 1931; it is the one from which the next-cited deductions are drawn; but as it does not differ markedly from its two predecessors, and is much more trying to the eyes than these, I do not reproduce it here. (In studying it, or the curve here pictured as Fig. 5, one must remember that the ordinate is proportional to the number of ions appearing per second per cubic centimetre, not of the atmosphere at the level in question, but of air of a standard density.)

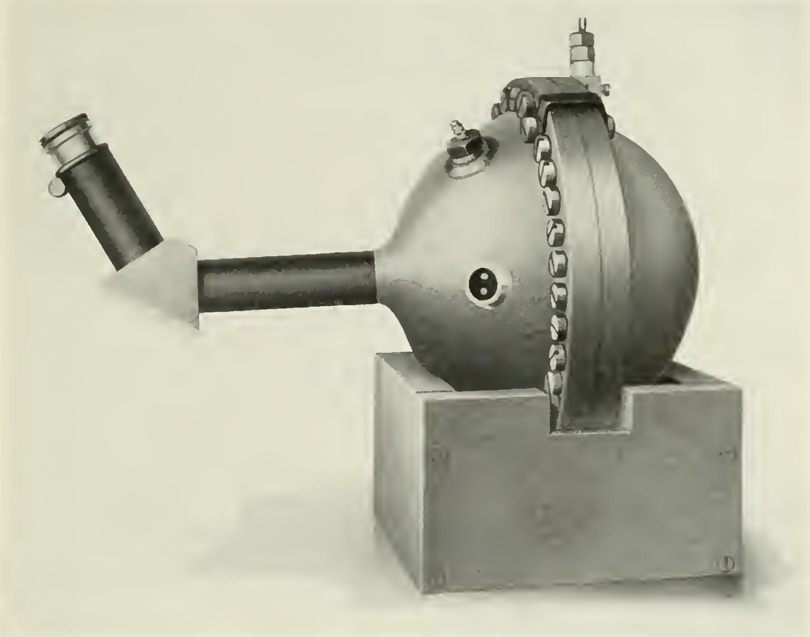


Fig. 6—Electroscope used in later under-water measurements of cosmic rays by Millikan and Cameron.

Is the whole of the curve, abstraction being made of the *Restgang*, an exponential curve such as would be found if the cosmic rays were a beam of gamma-rays of a single frequency descending vertically from above? The answer presumably is *no.* Millikan and Cameron, however, asked a slightly different question. Imagine that the cosmic rays, instead of coming altogether vertically from above, are gamma-rays traversing interstellar space uniformly in all directions, incident therefore upon the atmosphere at all angles between zero and 90°—much the most probable assumption if they do proceed from interstellar space, as it is hard to think of any possible action of the earth which might constrain them all to move straight toward its centre. The ionization at any depth x beneath the top of the atmosphere (continue to think of x as measured in metres of water) would then be due to beams coming from all directions above the horizon. These beams, having traversed different thicknesses of air, greater the more they are inclined to the vertical, would have been absorbed to different degrees; their intensities, originally (by hypothesis) the same, would be reduced in different proportions. It can readily be shown that the total intensity of all the beams should then vary, not as $e^{-\mu x}$, but as the following function of x:

$$y = C \int_1^\infty z^{-2} e^{-\mu x z} dz, \qquad (2)$$

μ standing as before for the absorption-coefficient of any cosmic-ray beam coming from any one direction.[6] Now the curves of Figs. 4 and 5, and the new curve of 1931, do not conform to this relation either. It is consequently not possible to affirm that the cosmic rays behave like gamma-rays of a single wave-length.

Does the final extremity of the Millikan curve conform with formula (2), so that one may assume that after a certain and feasible amount of filtering, the radiation is almost altogether reduced to gamma-rays of a single wave-length? To this Millikan and Cameron respond, that of the curve of Fig. 5 the portion extending from abscissa 30 to abscissa 60 *does* conform, the value required for μ being .0005. Of the curve of 1931 they say, that from 40 to 80 metres-of-water it conforms, if for the value .00028 be chosen.

Recalling Regener's value of μ (.00018) for depths still greater, and joining it with these, one is led to wonder whether at 40 and even at 80 metres the filtering is sensibly incomplete, and whether even at 200 metres the value ascribed to μ may not be the ultimate "hardness" of the rays. Further, there now arises the great question: can the curve as a whole be regarded as the sum of a limited number of terms of the type appearing in equation (2), with definite and distinctive values of μ? For if so, we may affirm that formally the cosmic rays behave like a mixture of gamma-rays having definite discrete frequencies, a sort of line-spectrum.

---

[6] See R. Hellmann, *Phys. ZS.* **30**, 357–360 (1929); E. Gold, *Proc. Roy. Soc.* **A 82**, 43–70 (1909). Regener used formula (1) instead of formula (2) for interpreting his data.

This is an extremely delicate question. What is needed is a method like that of Fourier analysis, whereby a given curve may be resolved into a sum, not of sines and cosines as result from the Fourier process, but of terms of the type of the right-hand member of (2); it seems that such a method is wanting. Millikan and Cameron, and independently their colleague Bowen, "built up . . . our observed curve out of four components—no smaller number would do—and in such a way that the synthetic and the observed curve [that of 1931] fitted exceedingly nicely from one to the other." These four components have the values .0080, .0020, .0010, .0003 of the coefficient μ; and their relative intensities, as inferred to exist at the confines of the atmosphere before the filtering commences, stand in the ratios 141000 : 130 : 80 : 33. It is evident that the portion of the radiation which survives at great depths of water, and for which the values of μ are extremely low, is a very small part of all that is to be found in the upper air. Yet even the highest of the values of μ here cited is much lower than the coefficient of absorption for the hardest known gamma-rays from radioactive substances.[7]

It would be deplorable to leave unmentioned some very romantic experiments in which the electroscope, with or without the observer, ascended by the aid of a balloon to heights of air hitherto unattained. Balloon-ascents were begun by German physicists (Gockel, Bergwitz, Hess, Kolhörster) in the five years before the war. Before, it had been maintained that the ionization of the air is due altogether to radioactive substances in the ground. The earliest balloon-flights impaired this argument, by proving that the amount of the ionization does not diminish rapidly as the observer flies upward; the later ones destroyed it, by proving that above a certain level (variously stated, and depending no doubt on the thickness of the walls of the chamber) the ionization rises with increase of height. I will however speak chiefly of more recent flights, those of Millikan's apparatus and that of Piccard.

The electroscope of Millikan and Bowen was borne aloft above the plains of Texas by a pair of balloons, one of which eventually burst and left the other to serve as a parachute in lowering its burden gently back to earth. In two experiments the electroscope climbed to heights of 11.2 and 15.5 kilometres, heights to which it seemed (until last year) that no man could possibly ascend and live. The apparatus (there were four of them) had to make its own records and bring them back; it was a masterpiece of ruggedness and compactness jointly, as one sees in Fig. 7, on the left of which it appears assembled (with a six-inch rule beside it to show its scale), on the right resolved into *disjecta membra*. The symbol *B* marks a manometer in the form of a U-tube with one arm closed, containing liquid; the symbol *T*, a thermometer in the form of a coil of metal with a pointer at its end. The light of the sky imprinted images of the meniscus of the barometer-liquid and the pointer of the thermometer-coil upon a moving film, kept in motion by the watch of which the mechanism is marked by *W*. Inside the cylinder *M* (so at least I read the text) were the two fibres of the electroscope already shown as Fig. 1, and they also were continually photographed by the light of the sky upon another film which the same watch kept in motion.

---

[7] Though not higher than that of the gamma-rays excited by Bothe and Becker when bombarding beryllium with alpha-particles, as I learn from a letter of Professor Bothe.

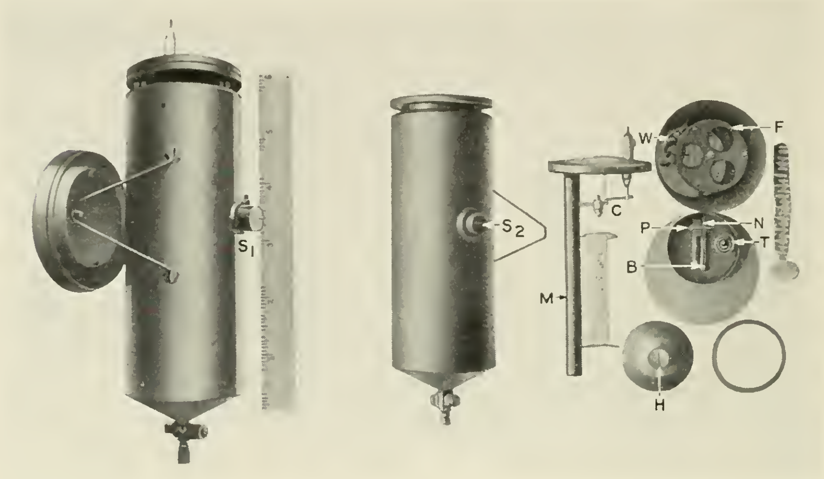


Fig. 7—Electroscope used in high-altitude measurements of cosmic rays by Millikan and Bowen (photographs somewhat retouched).

When one of these devices had ascended to 15.5 kilometres (nearly ten miles) and returned to earth eighty miles away from its starting-point, and had been picked up (by some casual wanderer, one infers) and returned to its authors, the films bearing the records of temperature and fibre-divergence looked as they do in Fig. 8. To determine the extent of the ionization at great heights without being deceived by the influence of temperature on the fibres, Millikan and Bowen compared the divergence of these when the balloon had reached a certain height (it was five kilometres) on the upward way, with their divergence as the apparatus passed the same level in descending. During the time while it was above this level, the fibres of the electroscope lost three times the charge that would have leaked away, if the instrument had been left on the ground. This again is proof that the cosmic rays are stronger in the upper air than in the lower.

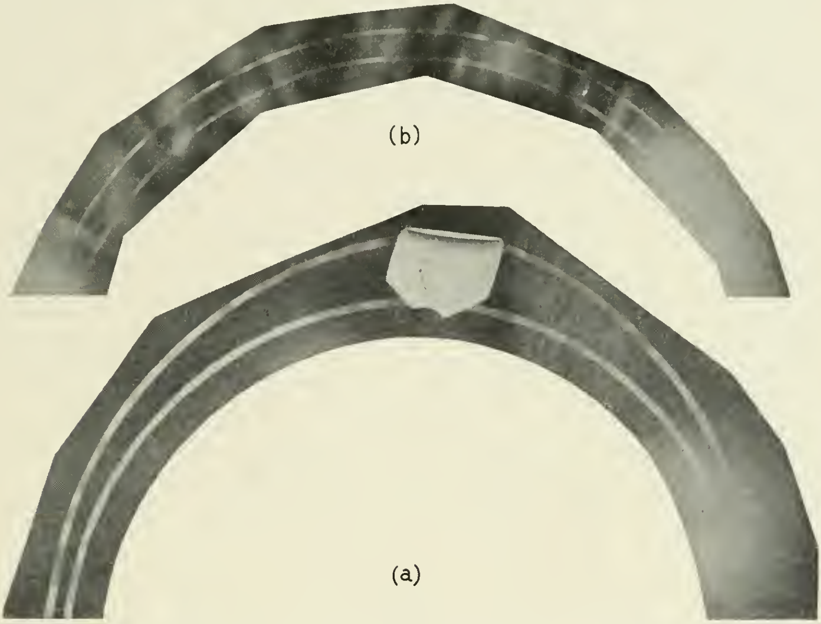


Fig. 8—Records obtained by automatic registration at high altitudes by Millikan and Bowen (somewhat retouched).

As I remarked above, until last year it had been deemed beyond the powers of living man to soar to heights so great as fifteen kilometres. This opinion was confuted by the superbly audacious flight of the Belgian physicist Piccard, which everyone will remember who reads the papers, though other things than cosmic rays were stressed in their accounts. Piccard seemingly has not yet published a full nor any extensive story of his measurements on cosmic rays, not at any rate in the physical journals; but in the bulletin of the French Physical Society I find an item in which he states as *provisional* result—awaiting further calibration of his instruments—that the ionization at a height of sixteen kilometres, in air at standard pressure, is of the order of 200 ions per cc. per second! This is considerably greater than the modest value of 1.5 prevailing at the level of the ground.

I must make at least a passing allusion to the *fluctuations* of the intensity of the cosmic rays—fluctuations which are smoothed over by some of the experiments and missed by others, for they are too small to be certainly detected by measurements of low precision, and many are too rapid to be noticed by measurements on the amount of discharge of an electroscope over a period so long as one or two hours.

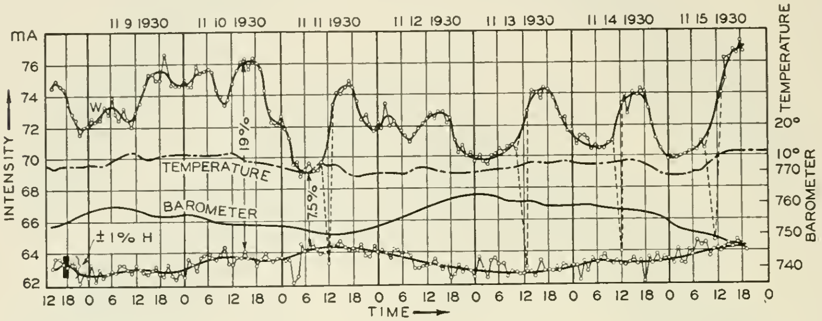


Fig. 9—Fluctuations of cosmic-ray intensity and attendant fluctuations of barometric pressure and temperature (Hoffmann).

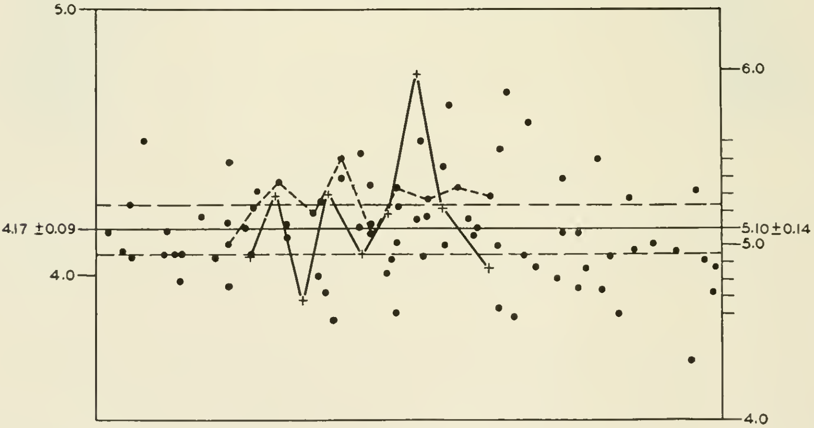


Fig. 10—Fluctuations of cosmic-ray intensity observed at high altitude in the Alps (von Salis).

Figs. 9 and 10 display examples of these: to appreciate the ratio of their average amount to the mean value of the ionization, one must take note of the scale of ordinates on the left. The former is especially instructive, for both the uppermost and the lowermost of the curves display the ionization of the air near the ground (in a basement at

Halle) measured at short intervals over the length of a week; the difference is, that the lowest shows the effect inside armor of lead ten cm. thick, while in getting the data for the topmost the plates of lead were removed from above the apparatus, though left in place at the sides and below. The lower curve thus represents the action of hard cosmic rays: one sees that it rises as the barometric pressure falls, falls as the latter rises—the heavier the blanket of air between Halle and the sky, the more these hard rays are reduced on their way to the ground. The upper shows the joint action of the hard cosmic rays, of others which are softer, and of soft rays from radioactive substances in the air: one sees how violently it fluctuates, supposedly (in Hoffmann's view) because the radioactive atoms wandering in the air change greatly in number as the weather changes. Figure 10 shows values of ionization obtained high up (3500 metres) in the Alps, on various August and September days of 1927, plotted against sidereal time; the zigzag lines connect observations made on two particular days. Corlin is said to have inferred, from a statistical study of data obtained in northern Sweden, that the ionization ascribed to cosmic rays decreases gradually and slowly before a magnetic storm, leaps suddenly to a high value at the onset of the storm, then decreases again.[8]

There is a related question which has been much debated: does the intensity of the cosmic rays vary as different celestial objects pass overhead, the sun for instance, or the Milky Way? As the years go on, the answer to this question becomes steadily more and more strongly in the negative. The opposite opinion has been held largely by German physicists; but a year ago one of them (Hess) reduced the proportion of the cosmic rays which he considers dependent on the sun, to half of one per cent of the total amount.

Certain perplexing data must be mentioned before we go on to the work which is done with other instruments than the ionization-chamber, for they impeach the reliability of this device. From many experiments of the pupils of Swann, it appears definitely established that if the quantity of air in an ionization-chamber is increased, the number of ions appearing in that chamber in unit time increases in a lesser ratio, and in fact approaches (and in some practical cases, even attains) a limiting value (Fig. 11). At first this seems natural enough: the limiting value should be attained when the air is so dense, that the ionizing rays are entirely absorbed before they have completely crossed the chamber. *But*, appreciable fractions of these ionizing rays are able to penetrate many metres of water, many centimetres of lead: it is pretty nearly a formal contradiction-in-terms, to assert that they can be fully absorbed in air of the density and thickness involved in these experiments. It seems necessary to suppose that as the pressure is increased, an ever-rising fraction of the total number of generated ions fails to make its way to the electrodes, positive ions and negative recombining with one another.[9] If so, there is reason to fear that the readings of a high-pressure ionization-chamber are affected by its temperature—a possible cause of spurious fluctuations.

---

[8] See W. M. H. Schulze, *Nature* **128**, 837–838 (14 Nov., 1931).

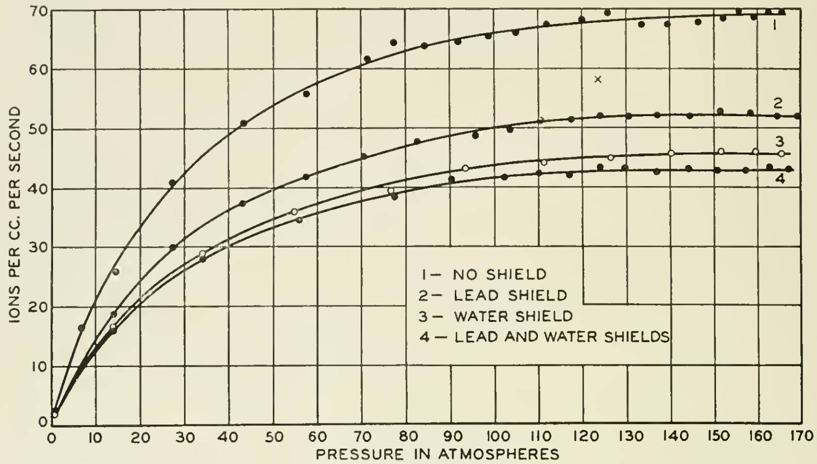


Fig. 11—Curves of ionization versus quantity of contained air in ionization-chamber (Broxon).

I turn now to the other ways of experimentation: there are two of them, both furnishing data more striking and spectacular than those of electroscopes. They exhibit the immediate cause of the mysterious ionization, and open the question of its ultimate cause. We must now take up this question: is the *ultimate* cause to be compared with electromagnetic waves or gamma-rays descending from above, as in this article I have thus far let be assumed? or, is it a rain of fast-flying material particles such as electrons or protons? or a mixture of the two classes of radiation? or has it properties unlike those of any radiation so far known?

(This question is often put in the words: are the cosmic rays wavelike or corpuscular? I think that this phrasing is very much to be condemned, gamma-rays as well as material particles being at one and the same time *both* wavelike *and* corpuscular. However the reader should know that in current usage "corpuscular" signifies "composed of particles of electricity or matter, instead of photons.")

---

[9] This theory, almost simultaneously published from Millikan's school and from A. H. Compton's, is at present under test.

The question is much more difficult to answer than it seems: for *even if the cosmic rays are of the nature of gamma-rays, the effect which they produce in ionization-chambers (and in the kinds of apparatus yet to be described) is almost entirely due to fast electrons.*

Consider a beam of X-rays (of the same nature as gamma-rays but softer) projected into a gas. It is known by experiment that they free electrons from atoms in two distinct ways: some of the corpuscles or photons are entirely absorbed in atoms, and tightly-bound electrons are ejected (photoelectric effect); some on the other hand make elastic impacts against loosely-bound electrons, communicating a part of their energy to these and retaining a part (Compton effect). As the frequency of the X-rays is increased, the former effect becomes less usual; entering into the range of gamma-ray frequencies, we find it fading out; it is reasonable to suppose that if there are corpuscles of light of frequency even higher than that of the gamma-rays, they ionize only by Compton-effect or "Compton scattering." It remains of course conceivable that such corpuscles may ionize in ways unknown at lower frequencies, as for instance by shattering the nuclei of atoms into fragments some of which are fast-flying electrons or protons.

Now though these impacts are properly called "elastic," their laws are peculiar, owing to the differences between corpuscles of light and corpuscles of matter.[10] The electron acquires some, usually a considerable part, of the energy of the photon; this latter (or as some would prefer to say, a new photon) goes off more or less obliquely, with lessened energy and lessened frequency. Given the wave-length of the light, the laws of the impact determine the energy acquired by the electron, the energy retained by the photon, and the direction of departure of the photon, as functions of the direction in which the electron is projected. If we know the distribution-in-direction of the projected electrons, we can determine their average energy. For electrons ejected by X-rays this distribution is known. But if the cosmic rays are (or include) gamma-rays whereof the absorption-coefficient has such values as those given (*e.g.*) by Millikan and by Regener for μ, they are much harder than any otherwise-known gamma-rays or X-rays, hence presumably of much lesser wave-length. Then, for the distribution-in-direction of the electrons which they eject, we must resort to an untested theory. This theory indicates that a beam of short-wave gamma-rays proceeding through matter should provide for itself an escort of electrons, most of them moving in directions inclined at very small angles to the beam itself, and many of them starting out with values of kinetic energy which are considerable fractions of the energy of the original photon.[11]

---

[10] See for instance my *Introduction to Contemporary Physics*, pp. 145–163, or the sixth article of this series.

It is this escort of fast electrons which should be the immediate cause of the ionization due to the short-wave gamma-rays. One easily sees how this result is destined to complicate the question of the nature of the cosmic rays. If it can be proved that the ionization attributed to these is caused by fast electrons, this may mean that the electrons are coming into the lower atmosphere from above, in which case most physicists would say "the cosmic rays are fast electrons." On the other hand it may mean that the electrons are produced in the lower atmosphere by photons coming from above, in which case most physicists would say "the cosmic rays are of the nature of gamma-rays." Yet in both cases the immediate agent would be the same; the term "cosmic ray" would be applied in the one case to it, and in the other to the ultimate cause of the immediate agent. This fine distinction, partly physical and partly verbal, is the source of a lot of confusion.

Consider now the evidence about the nature of this immediate agent. About three years ago the physicist Skobelzyn was engaged in studying the Compton scattering of gamma-rays from a certain radioactive substance, using the "expansion" or "cloud-chamber" method. By this method (it has often been described elsewhere,[12] I will therefore omit the details) the ions formed in a glass-walled container of gas are in effect rendered visible, each becoming the centre of a droplet of water, which appears on a photograph taken when the gas is illuminated. If during the experiment the chamber is traversed by a fast electron, its path is marked out by an unmistakable train of droplets. If a strong magnetic field is applied to the gas meanwhile, the trails of the fast electrons are visibly curved, and their speeds may be computed from the curvatures and the field strengths. Skobelzyn applied a field of 15000 gauss, in which the paths of the electrons liberated from atoms by the gamma-rays were curled up into beautiful spirals. But on examining the six hundred and thirteen photographs which he took, he found twenty-five trains of droplets resembling the curled ones due to the electrons of known cause, but not perceptibly curved at all! and two others of which the curvatures were perceptible, but so slight that out of them he estimated the enormous values of 7 and 15 millions of equivalent volts for the energy of the two particles. As for those of which the tracks were not sensibly curved by the field, the energy of some at least (assuming them to be electrons!) must have exceeded 15.10^6 equivalent volts. (Of the rest, the paths were so unfavorably placed that absence of sensible curvature might have been compatible with energy-values as low as, but no lower than, 3.10^6.)

---

[11] See for instance the article of H. Kulenkampff (*Phys. ZS.* **30**, 561–567; 1929) where he plots distribution-curves deduced by the quantum-mechanical theory of Klein and Nishina for electrons ejected by light of wave-lengths $2.4 \cdot 10^{-13}$ and $24 \cdot 10^{-13}$. He also considers the influence of the photons in continuing to eject new electrons after their first, second and later impacts; it appears that one photon is likely to start off several electrons at various points of the beam, instead of disappearing (so far as ionizing-power is concerned) after its first impact.

[12] As for instance in my *Introduction*, pp. 45–46, or the first article of this series.

This method has also been adopted by Millikan's collaborator, C. D. Anderson, who has a magnetic field of great extent, pervading a large expansion-chamber so oriented that particles coming along or near the vertical (we shall see evidence that near sea-level, the particles do favor that direction) are subjected over a long distance to its deflecting power, and electrons with energy values amounting to scores of millions of equivalent volts are sensibly deflectible. On some of his plates there are trails curved in the right sense for electrons coming downward from above, with energy amounting to 70 millions. On some there are trails curved in the opposite sense; if they were made by electrons, these must have been travelling from the earth upwards; if they were made by descending particles, as seems more plausible, these must have been positively-charged. On one there appear three paths, two apparently springing from a common origin near the wall of the chamber; one is curved in the proper sense for an electron, one in the opposite sense—if it is the track of a proton, this must have had energy of 120 millions—the third is sensibly straight.[13] Skobelzyn too had got plates on which two or three tracks appeared, coming probably from a common point of departure.

It appears from these pictures that the immediate agents of the ionization ascribed to "cosmic rays" are able to produce long trains of ions closely crowded together (the number of ions per centimetre is probably of the order of one hundred) [14] which is so far as our experience runs, a distinctive feature of electrified material particles such as electrons and protons; that some are deflected in practicable magnetic fields; that from the deflections it probably follows that the charge is sometimes negative and sometimes positive (the uncertainty being due to the fact that from the curvature one cannot tell the sign of the charge unless one knows in which sense the particle is going along the path); that the ones which are known to be charged have enormous energies, the others (those of which the paths are uncurved) either have yet more enormous energies or else are neutral. These facts make it impossible to suppose that photons are the immediate agents. Before continuing with the deductions we will consider the third method.

---

[13] I am much indebted to Dr. Millikan for showing and interpreting these plates to me. They have been shown in scientific meetings and mentioned in the press; the publication of the pictures and the work in the scientific journals will be eagerly awaited.

[14] The values most highly esteemed are obtained not by counting droplets, but by using the third method to determine the number of ionizing particles, the first to determine the total ionization; Kolhörster and Tuwim give 135 ± 10%.

The third method of studying cosmic rays involves the use of a "Geiger counter" or some modification thereof—a metal-walled gas-filled tube with a needle or (more commonly) a wire mounted inside it and insulated from the wall, which is connected to one pole of a battery, the needle or wire to the other. If the tube is properly designed and treated, the E.M.F. of the battery properly chosen (this seems to require a lot of experience) the passage of an ionizing particle across the gas between needle and wall is likely to evoke a sudden and violent and very short-lived current-flow. If a telephone-receiver is in the circuit, there is an audible click; if a galvanometer or electrometer is used with an optical device for recording its deflections upon a moving film, there will be a photographic trace of the discharge.

In experiments on cosmic rays with this device, it is the custom to employ a pair of counting-tubes (or even three) and to accept as valid data only the discharges which occur in either simultaneously with discharges in the other. I illustrate the method by a sketch from the work of Bothe and Kolhörster (Fig. 12) in which on the right one sees the two counters in longitudinal section, on the left one sees them in cross-section as they lie within their heavy armoring of metal.

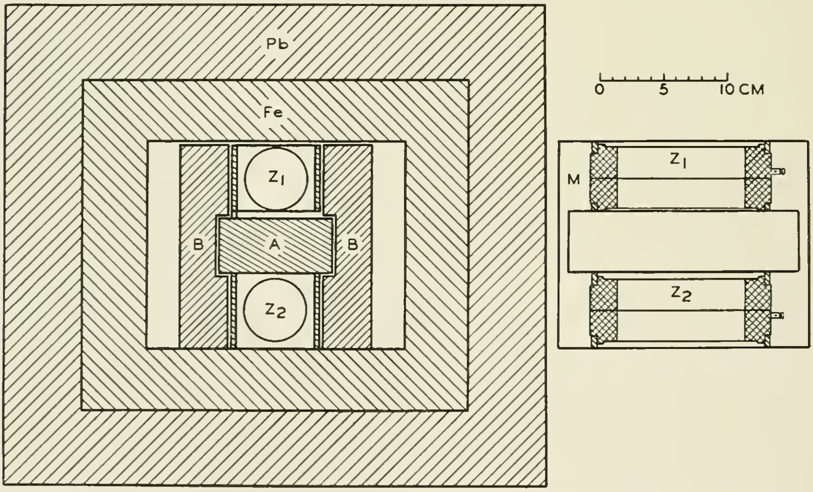


Fig. 12—Apparatus of Bothe and Kolhörster for counting ionizing particles and estimating the absorption thereof in metals.

"Coincidences"—which is to say, pairs of discharges occurring so close in time to one another that they seem to be simultaneous, that is, within an interval which with good apparatus may be a hundredth or even a thousandth of a second—are much more frequent than they would be, if they were due only to chance. Some of course *are* due to chance: these are the "spurious" or "casual" or "accidental" coincidences, of which the number must be estimated and subtracted from the number observed.[15] After this is done, one is pretty safe in assuming that the remainder is a measure of the number of ionizing particles which have darted through *both* of the counters. By this drastic procedure, of course, the observer limits himself to the study of those corpuscles which happen to be moving along paths which intersect both of the chambers. There will also be corpuscles of which the paths traverse one of the counters, but not the two; it seems as though the observer were throwing away his opportunity of studying these, by rejecting the non-coincident discharges; but some of the discharges in either tube are probably due to radioactive substances in its walls, or to unknown causes inherent in the tube and not connected with cosmic rays; and by accepting only the coincidences, one guards to a great extent against being misled by these. Moreover, by varying the relative position of the counters—for instance, by putting one of them first above and then beside the other—one may study the distribution-in-direction of the ionizing particles. Such studies have not yet been plentiful; but it has been found that if the counters are placed one above the other, the coincidences are several times more numerous than if they are placed side by side.

We have just been considering the effect of magnetic field on the tracks of the corpuscles ascribed to cosmic rays, observed by the cloud-chamber method: it is suitable now to review what the counting-tubes have to say about that question. Of the three physicists who have announced experiments upon the problem, L. M. Mott-Smith has given the most impressive account. There are three tubes, of the type exhibited in Fig. 12, their cylindrical chambers being some four cm. wide. Two are in the same vertical line, 30 cm. apart; the third moves to and fro along a horizontal line which intersects that vertical, 35 cm. below the lower of the two which are stationary. Triple coincidences are counted; they are most frequent when all three of the tubes are in line, as one would expect; when the movable tube is shifted away from the line, the number declines, the curve of coincidences N versus displacement x having "about the shape which is expected from the geometry of the arrangement, assuming rectilinear passage of the ionizing particles."

---

[15] Suppose that there are N_1 discharges per unit time in one of a pair of counters, N_2 per unit time in the other; and that if a discharge in either starts within a time T after the starting of a discharge in the other, the records of the two (whatever be the way of recording them) make them appear simultaneous. Then in unit time, the periods during which any discharge occurring in the first counter is not separately recorded add up altogether to an amount $2N_2T$. This is the fraction of time, discharges occurring during which are recorded as coincidences. If the events in the two counters are entirely uncorrelated, then out of N_1 discharges occurring per unit time in the first, $N_1(2N_2T)$ or $2N_1N_2T$ (on the average) will coincide with discharges in the second. If the discharges of each counter are recorded as peaks on tracings of separate films, one may determine N_1 and N_2 from the films, set for T the greatest time-difference which cannot surely be distinguished from zero and subtract $2N_1N_2T$ from the total number of coincidences. Or the number of accidental coincidences, together with N_1 and N_2, may be observed by placing the counters so far apart that no particle can traverse the two of them, and T thence computed. Or both sets of discharges may be recorded on the same tracing by connecting both counters to the same electrometer, in which case T will be the greatest interval for which two peaks merge apparently into one; or they may be so connected that no record whatever appears unless two discharges coincide.

There is another object in the vertical line beneath the lower of the stationary tubes, which I have not yet mentioned: it is a piece of iron 15 cm. thick, which in some of the experiments is unmagnetized, in some is magnetized in such a sense that charged particles which have descended vertically through the upper counters must be deflected, while remaining in that vertical plane which contains the horizontal line along which the movable tube is being shifted. If the particles are charged and of a single speed, the curve of N versus x should be shifted along the axis of x; if they are electrons of energy 10^9 (equivalent volts), it should be shifted by 2.2 cm., and should be easily distinguished from the curve obtained when the iron was demagnetized. Nevertheless, the two curves (or rather the three, for the iron was magnetized first in one and then in the opposite sense, in the hope of producing opposite deflections) are indistinguishable. Mott-Smith considers that if the ionizing particles be electrons, their energy must be not less than 2.10^9—a lower limit which may be halved, if they be protons. These values however are contingent upon an assumption, which I will mention in a moment.

The others who have entered upon this problem are Curtiss and Rossi. Rossi's first apparatus was remarkably like that which I have just described, independently and far apart though he and Mott-Smith were working—the one in Florence, the other in Texas. Later he modified the scheme, having two counters only, and a pair of slabs of iron so placed that when they are magnetized in one sense charged particles passing through the upper counter and thence into either slab should be deflected towards the lower counter, and when they are magnetized in the other sense the charged particles should be deviated away from the lower counter. Again the result was negative; and, says Rossi, "the corpuscular rays are not deflected . . . to such a great extent, as electrons or protons of 10 or 20 times 10^9 equivalent volts would be."

Unfortunately this, like Mott-Smith's, estimate is based upon a tentative answer to a question, one of the most perplexing in physics, though not so familiar as it should be to many who are interested in radiation, ionization of gases, and such phenomena: it is the ancient conundrum, *What is the strength of the magnetic field inside of a magnetized solid?*[16] Mott-Smith and Rossi both assume that the said field strength is equal to the induction B, admitting however that another assumption might be made. If instead one were to put it equal to the "magnetizing field" H, the lower limit inferred for the energy of the corpuscles would sink to a value so small, that the experiments would scarcely even be interesting. Considering this, it is especially important to recall the experiment of Curtiss which preceded the two others. He used two counters only, so foregoing the advantage of the narrow vertically-descending beam to which Mott-Smith confined his study; but between them he placed, not the iron of a magnet, but the space between the pole-pieces thereof; and he *did* get a positive result. The fieldstrength was 7000 gauss, the field extended over 24 cm. of the distance between the counters; vertically-descending electrons traversing the upper counter should have been deflected away from the lower, had their energy been 10^9 or less; actually "making allowance for accidental coincidences, a decrease of the order of 25 per cent has been observed."

We now take up a great but very thorny question: what do counters testify about the penetration and absorption of the cosmic rays in air, in water, and in metals?

Counting-tubes have been taken to great heights of air by Piccard, lowered to great depths of water by Regener. Neither has published (so far as I know) a full account of his data; but Regener implies that the curve of number-of-discharges-per-unit-time (it seems that one tube alone was used) versus depth-of-water agrees in shape with the curve of Fig. 3, especially in respect of its lowest end with the least value of μ. Bothe and Kolhörster took a pair of counters four hundred metres down into a mine, and found—if I read correctly—that there were no coincidences at all, apart from casual ones; a valuable result! The most striking work heretofore published was done in another way: by putting a thick piece of lead (or iron, or gold) either above or between a pair of counters set up in the same vertical line, and ascertaining in what proportion the coincidences are cut down.

The first who tried this out were Bothe and Kolhörster. In the particular experiment which I will quote, the apparatus of Fig. 12 was set up under a skylight on the top floor of a building not far above sea-level (the Reichsanstalt in Charlottenburg) *without* the upper part of the metal armor, so that rays coming vertically from above should reach the upper counter after passing through the atmosphere and a little glass. Between the upper and the lower counter, a block of gold 4.1 cm. thick was alternately inserted and removed. Over a period of twelve half-hours with the gold-block *absent* there were 19897 discharges in the upper counter, 12209 in the lower, 986 coincidences; over a period of twelve intercalated half-hours with the gold-block *present* there were 19814 discharges in the upper counter, 10562 in the lower, 743 coincidences. The presence of the gold had thus reduced the coincidences by 24.6 per cent; this the authors qualify by a considerable statistical uncertainty, stating it as (24.6 ± 4.2) per cent.

---

[16] I have discussed this question (without professing to solve it) in this journal, **6**, pp. 295–310 (1927). Or, see any textbook of magnetism.

Rossi did the like experiment, with lead and roofing over the upper counter, equivalent in thickness to 77 cm. of water; in place of the gold he used a slab of lead 9.7 cm. thick; it reduced the number of coincidences by sixteen per cent, which he expresses as (16 ± 3)%. Only a week or so before I add these words, he reported yet another.[17] There were three counters in a vertical line, 7 cm. of lead and 6 cm. of roofing above them, 25 cm. of lead permanently between them; and an additional thickness of *no less than seventy-six cm.* of lead, which he could interpose between the counting-tubes (half of it between the upper two, half between the two lower). This enormous absorbing layer cut down the number of (triple) coincidences by only some forty per cent (Rossi writes 38.5 ± 5.1).

Now, the coincidences are ascribed to ionizing particles such as those of which the tracks are made visible by the cloud-chamber method: particles which make densely-crowded trains of ion-pairs along their paths. Let us forget for the moment about high-frequency photons, and think only of such particles, descending almost vertically from on high. The numbers 24 and 16 and 38, quoted above, are the percentages of these which are stopped by plates of gold and lead, of the stated thicknesses, in the situations stated. Now the remarkable and perplexing point about these numbers is, that they are about the same as the percentage declines of ionization which are found, by Millikan and Regener, on lowering an electroscope through the equivalent thicknesses of water![18]

---

[17] I have just received (February 1) a preprint of the report from Dr. Rossi; it is marked *Naturwissenschaften* 20, 65 (1932).

[18] Bothe and Kolhörster, and Rossi, elect the following way of making the com-

At first glance this appears a simple and welcome result. Further study, however, dispels the pleasant impression: this is not an instance of the sort of observation which sometimes comes aptly to confirm a simple theory and smooth the path for further progress; it is instead a warning of dangerous complications. In introducing the curves of Figs. 3 and 4 and 5, I followed the policy of speaking of them as absorption-curves for rays analogous to gamma-rays descending from above; assuming in effect, that any ordinate of such a curve is a measure of the strength of the gamma-ray beam at the corresponding level of air and water. But if such rays there be, they ionize not directly, but by means of fast-flying electrons (or perchance, other particles) which they expel from atoms. A photon does not achieve the whole, nor even the most, of its ionization at the point where it dies; the effect is spread all along the path of the electron which is heir to a great part of its energy; and this path may be long. Thus, the ionization at a certain level of water or air may be due in part, it may even be chiefly due, to electrons which were unleashed at levels much higher. It may thus be chiefly due to photons which ended their careers far away from the place where it is being measured. It may then be no true measure of the number of photons which collide with electrons at the level in question, and no true measure of the strength of the gamma-ray beam at that level. Indeed it is not likely to be a true measure, unless the range R of the electrons be so short that the strength of the beam does not sensibly diminish as it descends through a distance equal to R. (There is an exception, as I mention on the next page; but in practice it may not help.) Now precisely this condition according to the testimony of Bothe and Kolhörster and Rossi, is *not* fulfilled.

Since the condition apparently is not fulfilled, all sorts of difficulties arise. If the ionizing particles have such ranges and such powers of penetration as the data suggest—not only the data just mentioned, but those of the cloud-chamber method and yet others—many of those which cause the ionization at any particular level of the air have come from points hundreds of metres, or even kilometres, farther up. Any ordinate of such a curve as that of Fig. 4 is a function of the whole of the (unknown) curve representing strength-of-gamma-ray-beam as

---

parison. Denote by I_0 the number of coincidences per unit time found with nothing between the counters, by I the number found with the block of gold (or lead) between, by d the number of grams per square cm. of gold (or lead). They equate I/I_0 to *exp* $(-\mu d)$, and compute the value of μ; in the first and the third of the experiments just mentioned, it comes out equal to $(3.5 \pm 0.5)\cdot 10^{-3}$ and $(1.6 \pm 0.3)\cdot 10^{-3}$; these are values of the same order of magnitude as those given by Millikan and by Regener for the quantity called μ in the analysis of their curves. In view of the objections to assuming exponential absorption-curves for rays which are or may be composed of fast-flying electrons, I doubt whether it is expedient to use this way of stating the results.

function of height, over an interval of this magnitude. This makes the problem of determining the trend of that curve (if such a curve there be) a formidable one.

Since the problem is so formidable and so important, let us linger over it. It is expedient to treat the very simplest conceivable case, a case which is certainly much simpler than reality, yet quite complicated enough for a first view. Suppose a beam of photons of a single wave-length descending vertically, first through vacuum, then (at $x = 0$) entering a horizontal sheet of some medium (water, say) of uniform density, in which they have an absorption-coefficient μ. Assume that every impact of a photon against an electron results in the total disappearance of the former, in the projection of the latter straight onward (vertically downward). Assume further that the projected electron engenders a constant number of ion-pairs per centimetre of its path, and that the length R of the path is the same for all the electrons.

Then, one readily sees that the ionization in the water, instead of being greatest at the top and diminishing steadily downward, actually increases from the top down to the depth R, and begins to decrease only beyond R. The formula is as follows, x standing for distance measured from the surface of the water downwards:

$$\begin{aligned} I &\propto 1 - e^{-\mu x} &\quad \text{for} \quad & x < R, \\ I &\propto e^{-\mu x}(e^{\mu R} - 1) &\quad \text{for} \quad & x > R. \end{aligned} \tag{3}$$

There is a "zone of transition" of thickness R, beyond which the ionization diminishes exponentially, with the same exponent as we have assumed for the absorption of the gamma-rays themselves. This is the exceptional case mentioned above, in which by diving to depths exceeding R one could arrive at a region where the trend of the ionization is the same as that of the strength of the beam of photons, and the value found for μ would apply to these last.

Unfortunately there is no actual case so simple. If there are cosmic gamma-rays coming from the sky, they evidently come from all directions, not merely vertically; and from the character of Millikan's curves it seems likely that they are of several or many wave-lengths, not one only. Conceivably at great depths of water the hardest of all may be filtered out, and these depths may be superior to the values of R for all the electrons; if so then the lowest values of μ, quoted by Millikan and by Regener, may pertain to actual photons. If there are cosmic gamma-rays of several frequencies descending vertically from above, then the actual ionization-curve should be a sum of terms such as (3), with numerical factors depending on the frequencies and the intensities, and values of μ depending on the frequencies; in principle, one might resolve the curve into these terms; but there is a great difference between "in principle" and "in practice." If instead of water we consider the atmosphere, there is reason to infer that the range R of the ionizing corpuscles discovered in cloud-chamber and counting-tube experiments may be greater than the thickness of the atmosphere itself! In this case, however, there should be a maximum of ionization lower down in the air, than the heights already attained by balloons; such a maximum has not been found. The possibilities are enormously complex; and after mentioning that Steinke has made a number of measurements with sheets of metal of thicknesses apparently inferior to the range of the electrons in the metal, with significant results (he speaks of *Uebergangseffekte* or "transitional effects"), I will pass to the description of an experiment by Rossi meant as a contribution to this problem.[19]

I have already quoted the value which Rossi obtained for the percentage drop in the number of coincidences occurring when a plate of lead is slipped between two counters—16 per 100. When he slipped the same lead plate above the upper counter, the drop was less—12 per 100. Let us imagine that photons coming from above were releasing fast forward-flying electrons from atoms in the plate of lead, these passing in the latter case through both of the counting-tubes, in the former through only one—causing coincidences therefore in the latter case, not in the former. The numerical values then oblige us to suppose that when 100 fast electrons dash from above into the upper counter, 16 of them can be stopped by ten centimetres of lead, but accompanying photons will generate four new ones in the slab, leaving a net decrease of 12 per 100. It is, then, this twelve per cent which is the decrease in the ionizing-power of the mixed beam, produced by interposing ten cm. of lead; Rossi thinks that the data prove that the photons are less penetrating than the electrons.

If so, how can it be that the charged particles responsible directly for the ionization have so great a penetrating-power? The electrons expelled from atoms by X-rays, and by the gamma-rays of known wave-length emerging from radioactive substances, are definitely less penetrating than the photons which set them on their ways. No theory gives the slightest indication that this relation should eventually be reversed as the frequency and penetrating power of the gamma-rays increases. True, the theories are based on the assumption that the energy of the ionizing particles is derived from the photons: if some of them originate from disintegrated atom-nuclei, as the seeming presence of protons among them suggests, the energy of these may be greater than that of the primary photon itself.

---

[19] I take equation (3) from Rossi's paper in the *Zeitschrift für Physik*, where the reader can also find equations for cases somewhat more complicated but probably not more plausible. The region beyond R is often called the region where "the primary beam is in equilibrium with its secondaries." Notice also the argument of Kulenkampff, mentioned in footnote 11. I am indebted to Dr. P. M. Morse for a discussion of these questions.

Could we dispense altogether with the notion of a beam of electromagnetic waves coming from above, and imagine that there are no cosmic rays other than these fast-flying particles, which then must be supposed to come down into the lower atmosphere from above? This possibility cannot be dismissed; against it, however, speaks the strong testimony of several observers who have travelled far and wide in the search for indications that the intensity of the rays varies from point to point on the surface of the earth. For the earth is a magnet; and while the strength of its field is minute compared with that prevailing within a few inches of an electromagnet or even an ordinary horseshoe magnet, the extent thereof is so great that electrons or protons coming up to our planet from interstellar space are liable to be enormously deflected. Charged particles of energy-values such as I have been mentioning (millions or tens of millions of equivalent volts) should reach the earth prevailingly in the region of the magnetic poles, if they start uniformly from all directions. Now, Bothe and Kolhörster took their instruments—three electroscopes and a pair of counting-tubes—on a cruise through Arctic waters; the ship started from Hamburg and returned there, its course encircling Iceland and passing close to Spitzbergen; there was no sign of a systematic variation of the readings. Millikan, whose data in California had agreed with his data in Bolivia in the admirable way which Fig. 4 displays, made a still finer test: in the summer of 1930 he went "to the settlement which is much the nearest to the earth's north magnetic pole of any settlement on earth, namely, Churchill, 750 miles due south of the pole on the west side of Hudson's Bay—at present a construction camp. . . . The mean results, when compared with those similarly taken at Pasadena during the last week in July and the first in August, show that the cosmic rays have precisely the same intensity at Churchill, in latitude 59, as at Pasadena in latitude 34, the mean results in the two places being 28.31 ions per cc. per sec. and 28.30 ions per cc. per sec. respectively, as measured in my particular electroscope. I think the error in these measurements cannot possibly be as much as 1 per cent." A British expedition to the Antarctic made measurements of cosmic rays within 250 miles of the south magnetic pole, and found the same intensity there as in Australia.[20] It is difficult to doubt that anything so constant must come from without the terrestrial world, and on its way be exempt from the influence of the earth's magnetic field.[21]

---

[20] K. Grant, *Nature*, **127**, 924 (1931).

Of the origin of the cosmic rays I have not spoken in this article, having so chosen its title as to exclude that mighty subject from its scope. In the data and in the nature of these rays there is enough of the mysterious and enough of the extraordinary, to suffice for an introduction. Modest as the data seem, what they reveal is sensational. It is found that the atmosphere of this earth is being traversed by ionizing particles, of which the qualities are amazing. If they are such familiar corpuscles as protons or electrons, their energy must be of the order of tens of millions of equivalent volts, values without precedent in our experience. If they are neutral particles they are in themselves unprecedented. If they are electrons or protons which derive their energy from photons or corpuscles of light, these last must have energy greater, frequency higher and wave-length smaller, than any form of light of which we have previous knowledge. Such quantities of energy, be it remembered, are enormously greater than the largest which atoms can emit in the course of their normal lives; they are several times larger even than those which are emitted by collapsing atom-nuclei, in those processes of transmutation of which we already have knowledge. If they come from individual atoms they must arise from processes heretofore unknown, of transmutation or synthesis or annihilation. The constancy of the effect of the ionizing rays, its independence (be it absolute or only approximate) of weather and time and direction and the earth's magnetic field, implies that these processes are spread, not over the earth nor even over the solar system by itself, but throughout the whole of the cosmos.

LITERATURE

The literature of cosmic-ray research has already swollen to such proportions, that a complete bibliography would cover several pages of this journal. There seems to be no book devoted altogether to the subject, though in treatises on atmospheric electricity—I name especially *Die elektrische Leitfähigkeit der Atmosphäre*, by V. F. Hess, translated as *The Electrical Conductivity of the Atmosphere and its Causes* (Van Nostrand, 1926)—a chapter or a section is usually assigned to it.

The major publications of Millikan's school have appeared in the *Physical Review*, as follows:

R. A. Millikan & I. S. Bowen, *Phys. Rev.* (2), **27**, 353–361 (1926) (sounding-balloon observations).
Millikan & R. M. Otis, *ibid.*, **27**, 645–658 (1926) (mountain-peak and aeroplane observations).
Millikan & G. H. Cameron, *ibid.*, **28**, 851–868 (1926) (lakes in California).
Millikan & Cameron, *ibid.*, **31**, 163–173 (1928) and 921–930 (lakes in California and Bolivia).

---

[21] Nevertheless there have been opposing data; Clay, on a voyage to Java, from the Mediterranean, found the ionization increasing as he approached the equator, and Corlin is said to have observed an increase from south to north along the Scandinavian peninsula; recall also Corlin's assertion mentioned on page 169.

Millikan & Cameron, *ibid.*, **32**, 533–557 (1928) (interpretation).
Millikan, *ibid.*, **36**, 1595–1603 (1930) (approach to north magnetic pole).
Millikan & Cameron, *ibid.*, **37**, 235–252 (1931) (most recent and complete ionization-*vs.*-depth curve).

To this list should be added Millikan's address before the B. A. Symposium on the Evolution of the Universe (*Nature*, **128**, 709–715; 24 Oct. 1931), and the account of C. D. Anderson's work probably soon to be published.

Regener's observations at various depths of water are described in *Naturwissenschaften*, **17**, 183–185 (1929), with a later and regrettably brief note in *Phys. ZS.*, **31**, 1018–1019 (1930). Piccard's observation at the altitude of 16 km. is mentioned in the *Bulletin* of the French Physical Society for 4 Dec. 1931. Swann's observations at three altitudes appear in *Jour. Franklin Inst.*, **209**, 151–200 (1930), and an observation by A. H. Compton in *Phys. Rev.* (2) **39**, 55 (1932).

Skobelzyn's experiments by the expansion-chamber method are published in *ZS. f. Phys.*, **54**, 686–702 (1929); see also P. Auger & D. Skobelzyn, *C. R.*, **189**, 55–57 (1929). L. M. Mott-Smith and G. L. Locher made simultaneous observations by this and the Geiger-counter method (*Phys. Rev.* (2), **38**, 1399–1408; 1931).

The absorption-experiments with Geiger counters cited in this article are due to W. Bothe & W. Kolhörster, *ZS. f. Phys.*, **56**, 751–778 (1929), and B. Rossi, *ibid.*, **68**, 64–84 (1931); see also Rossi, *Lincei Rendiconti* (6), **13**, 600–606 (1931).

Experiments with Geiger counters to seek for deflection of ionizing particles by magnetic fields: Mott-Smith, *Phys. Rev.* (2), **37**, 1001–1003 (1931); Rossi, *Lincei Rendiconti* (6), **11**, 478–482 (1930), and *Nature*, **128**, 300–301 (22 Aug. 1931); L. F. Curtiss, *Phys. Rev.* (2), **35**, 1433; (1930).

Search for influence of earth's magnetism: Millikan, *ll. cc.;* K. Grant, *Nature*, **127**, 924 (20 June 1931) (approach to south magnetic pole); Bothe & Kolhörster, *Berl. Ber.* (1930) (cruise in Arctic waters); J. Clay, *Proc. Amsterdam Acad.*, **30**, 1115–1127 (1927), and **31**, 1091–1097 (1928) (cruise to tropics); Rossi, *Nuovo Cimento*, **8**, 3–15 (1931); W. M. H. Schulze, *Nature*, **128**, 837–838 (14 Nov. 1931) (magnetic storms).

On the dependence of cosmic-ray ionization on time there have been many researches, mostly by people desirous of finding (or of *not* finding) influences of barometric pressure, weather, variations of the earth's magnetic field, the position of the sun, the position of the Milky Way or other remote bodies. Most if not all of the attempts to determine the absolute value of the ionization to be ascribed to cosmic rays, on or above the ground, have involved by necessity studies of fluctuations. Fig. 9 of this article comes from G. Hoffmann, *ZS. f. Phys.*, **69**, 703–718 (1931); his paper contains references to earlier German work, some of it in a geophysical journal. Fig. 10 above comes from G. von Salis, *ZS. f. Phys.*, **50**, 793–807 (1928). Millikan discusses the fluctuations in his 1930 paper. The references to Hess' estimate of the influence of the sun are: *Nature*, **127**, 10–11 (3 Jan. 1931) and *ZS. f. Phys.*, **71**, 171–178 (1931); earlier work is mentioned there.

The estimate made by Kolhörster and Tuwim of the average number of ion-pairs produced (in air under standard conditions) per centimetre path of one of the high-speed ionizing particles detected by the Geiger counters, is stated in *ZS. f. Phys.*, **73**, 130–136 (1931); for the notation it seems to be necessary to study an earlier paper in *Berliner Berichte*, 1931.

For effects associated with the "transition-zone" mentioned above on page 180, see E. Steinke, *Phys. ZS.*, **31**, 1019–1022 (1930) and the earlier work there cited; A. Corlin, *ibid.*, 1065–1071.

For the variation of the ionization inside a high-pressure air-chamber with the pressure of the air, see J. W. Broxon, *Phys. Rev.* (2), **37**, 1320–1337 (1931) and the literature there cited; E. G. Steinke & H. Schindler, *Naturwiss.* **20**, 15–16 (1932) and literature there cited. For the theory mentioned on page 170 *supra*, see Millikan & Bowen, *Nature*, **128**, 582–583 (1931); A. H. Compton, R. D. Bennett & J. C. Stearns, *Phys. Rev.* (2) **38**, 1565–1566 (1931); Steinke & Schindler, *loc. cit.*

Abstracts of Technical Articles from Bell System Sources

The Emission of Secondary Electrons from Tungsten.[1] A. J. Ahearn. An apparatus is described for investigating critical potentials in the emission of secondary electrons from tungsten. Measurements on the velocity distribution in the primary beam show that secondary electrons from the electron gun are absent. Tube characteristics which might introduce spurious critical potentials in the secondary emission from tungsten appear to be absent. By heat treating the tungsten and cleaning up residual gases, maxima and critical slope changes were developed below 40 volts. With sensitive methods of measuring and plotting the data, critical potentials within the range from 40 to 500 volts were observed only at the following uncorrected voltages: 70, 108, 208, 297 volts. All but the 70-volt effect disappeared eventually after heat treatments of the tungsten target. Thus when the tungsten surface is most free from contamination, critical potentials persisted only at the following uncorrected voltages: maxima at about 3.5 and 8 volts and slope increases at 24, 33 and 70 volts. The phenomena may be associated with the diffraction of electrons or the production and absorption of characteristic soft x-rays. Regardless of the mechanism operating at the critical potentials, their decrease or elimination beyond 40 volts points strongly to effects of surface contamination rather than to characteristics of tungsten.

Electrolytic Phenomena in Oxide Coated Filaments.[2] J. A. Becker. A critical survey of the literature shows that the current through the oxides in oxide coated filaments is carried by electrons, negative oxygen ions, and positive barium ions. The proportion of current carried by each depends upon the exact composition and method of preparation of the oxide coating, on the heat treatment and on previous electrolytic effects. Presumably the conductivity is greatly affected by barium and oxygen dispersed through the oxide. New experimental results show:

1. For a particular BaO + SrO filament, the conductivity C was given by

$$1.71 \times 10^{4} \frac{1.73 \times 10^{4}}{T} + 5.55 \times 10^{-3} \frac{0.62 \times 10^{4}}{T} \cdot$$

---

[1] *Phys. Rev.*, November 15, 1931.

[2] *Trans. Electrochem. Soc.*, Vol. LIX, 1931.

2. The current is proportional to the voltage only so long as the current is small; otherwise the products of electrolysis alter the conductivity.

3. Polarization currents are caused by the Ba and O which are produced by electrolysis. These currents decrease rapidly even at temperatures near 500° K., thus showing that Ba and O diffuse at low temperatures.

Some Observations of the Behavior of Earth Currents and their Correlation with Magnetic Disturbances and Radio Transmission.[3] Isabel S. Bemis. This paper presents correlations between the abnormal earth currents noted during magnetic storms and transoceanic radio transmission on both long and short waves. The radio transmission data were collected on the telephone circuits operating between New York and London and between New York and Buenos Aires. The earth current data were collected on two Bell System lines extending approximately a hundred miles north and west from New York. The results of this work establish facts which have been known in a general way for some time.

The direction of flow of abnormal earth currents in the neighborhood of New York seems to be along a northwest-southeast line. Coincident with such abnormal currents are periods of poor short-wave radio transmission. However, on long waves, daylight transmission over transatlantic distances is improved. On the short-wave circuit to Buenos Aires, transmission is adversely affected but only to a moderate extent.

The Propagation of Short Radio Waves over the North Atlantic.[4] C. R. Burrows. Transmission conditions for each season are shown by "surfaces" giving the received field strength as a function of time of day and frequency. These show that frequencies near 18 mc are best for daytime transmission. In summer the best frequencies for nighttime transmission are those near 9 mc. In winter an additional frequency near 6 mc is required during the middle of the night. A frequency (such as 14 mc) intermediate between the day and night frequency is useful during the transition period between total daylight and total darkness over the path. Day-to-day variations change the periods of usefulness of these frequencies. In particular the period of usefulness on 14 mc sometimes extends so that it is the best daytime frequency.

Transmission conditions on undisturbed days were found to be the same for the same time of year on different years. These undisturbed transmission conditions are presented by "normal" surfaces. Comparison of these surfaces shows that the higher frequencies are less attenuated in winter. Reception on the highest frequency, 27 mc was best in winter; in summer this frequency was never heard.

---

[3] *Proc. I. R. E.*, November, 1931.

[4] *Proc. I. R. E.*, September, 1931.

The effect of solar disturbances on short-wave transmission is to reduce reception on all frequencies. Sometimes the higher frequencies are the more adversely affected. Some of the possible causes of these disturbances are discussed.

From the measurements made on "static" at New Southgate, data on the variation of its field strength as a function of frequency, time of day, and season are given.

Methods of High Temperature Treatment.[5] PAUL P. CIOFFI. The object of this paper is to describe methods which have been developed in this laboratory for treating metals, chiefly iron and its alloys, at all temperatures up to about 1700° C., considerably above the melting point of iron, and in any atmosphere ranging in pressure from 10^{-8} mm. of mercury to 20 or more atmospheres. The methods of heating, and the forms of the materials heated fall into three rather well-defined groups: (1) Long wires and tapes heated to any temperature up to a few degrees below the melting point by passing currents through them in the presence of a gas (*a*) under a pressure of one atmosphere or less, or (*b*) under a pressure up to 20 atmospheres. (2) Toroids heated by induction in a gas atmosphere, with the gas pressure ranging from 10^{-8} mm. to one atmosphere. (3) Any shape of specimen heated up to about 1700° C. in a molybdenum wound furnace in a gas pressure ranging from 10^{-3} mm. to 20 atmospheres.

In investigations dealing with the effect of heat treatment on the properties of materials, the temperature range covered, often limited by the lack of suitable facilities, is likely to obscure important or interesting effects produced by heat treatments at high temperatures. The new high values of the magnetic permeability of iron recently reported were made possible by employing the methods here described; in fact, these methods were all developed in connection with the investigation of the magnetic properties of materials as dependent upon temperature and time of heat treatment and rate of cooling in any atmosphere. These methods are also applicable to other metals which can be shaped into long wires or tapes or toroids.

---

[5] *Jour. Franklin Institute*, November, 1931.

Methods for Measuring Interfering Noises.[6] LLOYD ESPENSCHIED. This paper outlines various methods of measuring interference, particularly in radio telephony, which have been found useful in the Bell System.

Reverberation Time Measurements in Coupled Rooms.[7] CARL F. EYRING. The decay of residual sound in simple reverberant enclosures has been the subject of much study. Out of these investigations along with other important developments formulæ have emerged which may be used to calculate the reverberation time of enclosed rooms, the one proposed by Sabine being applicable to live rooms, but a more general one being necessary if applied to dead rooms. These formulæ may be used in the study of simple enclosures in which the sound is diffuse, the absorbing material is well distributed, and the decay of the residual sound is exponential in time. They may not be applied indiscriminately to complex structures for the author has shown that the curves illustrating the decay of the residual sound intensity level may not be straight under certain conditions for coupled rooms of different natural reverberation times or even for a single room with no sound diffusing scheme and non-uniformly distributed absorption.

The present paper presents further data on acoustically coupled rooms and offers a theoretical study and formulæ applicable to such complex structures. Somewhat similar studies have been made by Buckingham and Davis in the investigation of sound transmission through partitions. But before the coupled room formulæ which are based on certain idealizations can be applied in general to auditoriums with various types of balconies and under-balcony spaces, they must be carefully checked by a thorough experimental study of typical theatres. Thus it is expected that empirical formulæ, based on theory and experiment, may be evolved for each type of complex structure.

Recently developed instrumental methods of measuring reverberation time, especially those methods that measure the decay history of the residual sound, give promise of being the tools needed in this study. The results recorded in this paper were made on a meter described by Wente and Bedell in their paper "Chronographic Method of Measuring Reverberation Time" and by the author. This instrument plots almost automatically the intensity level of the residual sound measured in db and the time.

[6] *Proc. I. R. E.*, November, 1931.

[7] *Jour. Acous. Soc. Amer.*, October, 1931.

When the sound source is cut off there will not only be a decay of sound due to the absorption of the walls, fixtures, etc., but the interference pattern will continually shift causing the actual rate of decay of the sound intensity level at a point to fluctuate about the rate of decay caused only by absorption. This fluctuation may be very pronounced and actually changes from point to point in the room, but may be minimized by the use of a warble tone and if necessary by the proper placement of the transmitter. With these precautions, and they are used in all the reverberation time measurements presented in this paper, the effect of interference is greatly reduced. The average rate or rates of decay δ (db per second) of the sound intensity level can at once be obtained from the slope of the straight line or lines which best fit the series of experimental points. The reverberation times T (time per 60 db) used in this paper have been calculated using the relation

$$T = \frac{60}{\delta}.$$

The Effect of Exposure and Development on the Quality of Variable Width Photographic Sound Recording.[8] DONALD FOSTER. This paper deals with the dependence of the quality of variable width recording on the conditions of exposure and development. When the widths of the images employed in recording or reproducing are comparable with the wave-length of the record the exposed portion of the record is not uniformly exposed. The record is attenuated in amplitude as the frequency is increased, and harmonics are introduced whose relative intensities depend on the contrast of development and on the frequency. When the exposure of the record occupies the linear range of the H & D curve, and when the product of the gammas of the negative and the positive is equal to unity, the record is practically free from spurious harmonics. The amount of non-linear distortion is calculated for the case when the over-all gamma is equal to two; and it is shown that Cook's analysis of the aperture effect gives a superior limit to the distortion obtainable by overexposure or by over or underdevelopment. The effect of the unavoidable non-uniform illumination of the images is considered.

The Vectorial Photoelectric Effect in Thin Films of Alkali Metals.[9] HERBERT E. IVES. It is assumed that the photoelectric effect exhibited by thin films of alkali metals on specular platinum surfaces is proportional at any wave-length to the electric intensity just above

[8] *Jour. S. M. P. E.*, November, 1931.

[9] *Phys. Rev.*, September 15, 1931.

the platinum. This electric intensity is found, using the optical constants of platinum, by computing the intensities of the wave patterns formed by the interference of the reflected and incident beams. These computations are made for various angles of incidence and for light polarized in and at right angles to the plane of incidence. The intensities thus found exhibit very large ratios of value for the two planes of polarization, in striking agreement with the characteristics of the vectorial photoelectric effect. The changes of amplitude of the perpendicular electric vector on entering the alkali metal film, as computed from the optical constants of the alkali metal, account for the experimentally found low values of the emission ratios at long, and their high values, at short wave-lengths.

The Photoelectric Effect from Thin Films of Alkali Metal on Silver.[10] HERBERT E. IVES and H. B. BRIGGS. The thin films of alkali metals which spontaneously deposit in vacuo on other metals have long been known to exhibit photoelectric effects which vary in amount and character, depending on the underlying material, but the exact nature of this dependence has been obscure. Silver, because of its region of exceedingly low reflecting power in the ultraviolet and the accompanying variation of optical constants, is exceptionally well suited for studying the influence of the underlying metal. It is found that the region of low reflecting power profoundly affects the photoemission, but in a manner not to be explained simply by reduction of light reflected back through the alkali metal film or by the absorption of light by the silver. The results obtained are very satisfactorily explained upon computing, from the optical constants, the intensity at the surface, of the interference pattern formed by reflection just above the silver surface. The positions of the maxima and minima of photoemission, and their variations with angle of illumination and plane of polarization are accurately indicated.

The Applicability of Photoelectric Cells to Colorimetry.[11] HERBERT E. IVES and E. F. KINGSBURY. It is the purpose of this paper to consider critically the requirements for a precision physical colorimeter, and to estimate, in the light of a large body of experimental data on the new types of photoelectric cells, to what degree the requirements for physical colorimetry may be met at the present time. The paper is intended to be very specifically limited to the problems of precision color measurement, and it is assumed that the reader is already acquainted with the principal facts relating to photoelectric cells and their use in ordinary photometric measurements, in which no pronounced color differences are faced. Our study does not concern itself with certain problems of sorting and selection, commonly spoken of as "color" measurement, in which the nature of the colors concerned and their range of deviation from certain standards are known in advance, thereby greatly simplifying the problem.

---

[10] *Phys. Rev.*, October 15, 1931.

[11] *Jour. Optical Soc. America*, September, 1931.

A Moving Coil Microphone for High Quality Sound Reproduction.[12] W. C. Jones and L. W. Giles. A microphone is described in this paper which retains all of the inherent advantages of the moving-coil type of structure but unlike the earlier forms of this microphone responds uniformly to a wide range of frequencies. It is more efficient than the conventional form of condenser microphone and its transmission characteristics are unaffected by the changes in temperature, humidity and barometric pressure encountered in its use. Unlike the condenser microphone the moving-coil microphone may be set up at a distance from the associated amplifier and efficient operation obtained. Owing to its higher efficiency and lower impedance it is less subject to interference from nearby circuits. It is of rugged construction and when used in exposed positions is less subject to wind noise.

The Shot Effect in Photoelectric Currents.[13] B. A. Kingsbury. The shot effect, as it occurs in a photoelectric current, has been used to secure an evaluation of the electron charge. A new and original method of amplifier calibration, which involved the use of a modulated light beam, simplified the measurements and the computation of the result. In the absence of space charge, the experimental value of the electron charge was 1.61×10^{-19} coulombs for a thermionic current, and about 25 per cent greater for a photoelectric current. It was found that the shot effect is enormously increased in photoelectric currents which are amplified by collision ionization. Statistical variations which might be expected to occur in a beam of radiant energy could not be detected, since, within the limits of experimental accuracy, the shot effect in photoelectric currents was found to be independent of the frequency of the light producing electron emission.

Some Acoustical Problems of Sound Picture Engineering.[14] W. A. MacNair. The purpose of this paper is to point out that many advances in acoustical engineering have been necessary in order to understand and control adequately the conditions under which modern sound pictures are recorded and reproduced. To illustrate this point, some of the acoustical problems encountered at Bell Telephone Laboratories are discussed. The sudden and successive changes in sound intensity level to be expected in a room during the growth and decay of sound from an intermittent source are pointed out. The necessity of using the more general reverberation time formula, which was developed over a year ago, when dealing with comparatively "dead" rooms, is indicated. One type of acoustical distortion which is due to interference is discussed together with the measures necessary to minimize it in sound pick-up work. These phases of acoustical engineering have been selected for discussion from many which confront the engineer in this field.

---

[12] *Projection Engineering,* October, 1931.
[13] *Phys. Rev.,* October 15, 1931.
[14] *Proc. I. R. E.,* September, 1931.

An Interpretation of the Selective Photoelectric Effect from Two-Component Cathodes.[15] A. R. Olpin. Evidence is produced to support the view that photoelectrically selective, two component cathodic surfaces are crystalline in nature. Then, assuming that Fowler's equation for the energy of electrons selectively transmitted through a single potential valley $[W = (n^2h^2/8md^2)]$ is equally valid for the energy of electrons selectively transmitted through the periodic sequence of valleys characteristic of the potential field within a crystal, and that all of the energy of photoelectrons is acquired from the incident light quanta, the wave-lengths of light to which such a surface should respond selectively can be computed. Such computations have been made with d equal to the internuclear distance between electro-positive ions in the lattice structure of alkali metal hydride, oxide and sulphide crystals. The hydride crystals belong to the sodium chloride type and the oxide and sulphide crystals are supposedly of the calcium-fluoride type. The correlation between these computed values and the positions of the observed selective maxima is exceptionally good. Moreover, the fact that the alkali metal hydrides exhibit but one selective maximum and the oxides two or three maxima is in keeping with the geometry of their respective crystalline types.

Some Physical Concepts in Theories of Plastic Flow.[16] R. L. Peek, Jr., and D. A. McLean. A review is given of the considerations involved in the development of theories of deformation applicable to the flow of soft solids in capillary tubes and under similar steady state conditions. It is pointed out that the limitation to special test conditions (particularly to steady states) makes it impossible to distinguish in all cases between the results of different physical hypotheses as to the mechanism of flow resistance. A review is given of the physical distinctions which can be determined by such experiments. It is noted that the criterion that the results from different capillary tubes can be expressed in the form $Q/R^3 = F(PR/2L)$ serves to distinguish those types of flow in which the resistance is dependent only on the rate and not on the extent of the deformation, except when slip occurs. This last can be distinguished from quasi-laminar flow (in which the resistance depends on the amount of deformation) by tests made with capillaries of common radius but of different lengths.

---

[15] *Phys. Rev.*, November 1, 1931.

[16] *Jour. Rheology*, October, 1931.

The other important physical distinction that can be observed is between those materials which show a yield value and those that do not. For those that do, a new type of equation is obtained, of which the Bingham-Buckingham equation is a special case. In this new general form the relation postulates an initial shear stress which must be exceeded before flow takes place, and a lower constant stress which is effective in opposing flow once the latter has commenced.

For cases of flow in which no yield value is observed, another new form is given which is based on the physical concept, common to most theories of such flow, of an effective viscosity varying with the stress intensity between upper and lower limits. It is shown that the character of this relation is similar to that given by the empirical formula $Q/R^3 = K\left(\frac{PR}{2L}\right)^n$. The use of this equation in expressing and interpreting experimental results is discussed.

Intercontinental Radiotelephone Service from the United States.[17] J. J. PILLIOD. Radiotelephone service between the United States and Europe was established January 7, 1927 with one circuit and with service to limited areas. Facilities and service have been greatly improved and extended and rates have been reduced. Present scope of service is described and reference made to consistent increases in transatlantic telephone messages handled. This increase indicates that this service is being found of increasing value by the public.

Extent of ship-to-shore radiotelephone service from the United States is outlined. Arrangements for service to Buenos Aires and Rio de Janeiro are described, these differing from arrangements used for service to Europe in that operation to these two cities was planned on a part time basis. Proposed short-wave system for operation with

---

[17] Presented at Pacific Coast Convention of A. I. E. E., Lake Tahoe, Calif., August, 1931. Published in abridged form in Elec. Engg., September, 1931.

Bermuda and proposed new long-wave system to supplement existing facilities to Europe are mentioned.

A description of the new radiotelephone transmitting and receiving stations now being erected at Dixon and Point Reyes, Calif., respectively, is given. These stations will be connected to a terminal office at San Francisco and the system used for the establishment of radiotelephone service to the Hawaiian Islands and later on, to other transpacific points as may be required.

High-Frequency Atmospheric Noise.[18] R. K. POTTER. A method which has been employed in the measurement of high-frequency atmospheric noise is described. Using this method measurements of noise over the range from 5 to 20 megacycles made in different parts of the United States and at different times of the year, show a distinct diurnal change in intensity similar to that for long-range high-frequency signal transmission. Except during periods of severe local disturbance noise on the lower frequencies is high during the night while on the higher frequencies the maximum occurs during the day. Simultaneous observation of crashes on different frequencies also suggests that the received atmospherics are largely transmitted by overhead paths. The variation in high-frequency atmospheric noise intensity during the passage of local electrical disturbance centers is shown. It is suggested that the intensity of atmospheric noise generated by these centers of electrical disturbance is inversely proportional to frequency. Measurement data are included showing the effect of sunrise and sunset, an eclipse of the sun, and disturbances in the earth's magnetic field upon the intensity of high-frequency atmospheric noise. Diurnal characteristics of high-frequency atmospheric noise on directive antennas facing England and South America and the noise reduction obtained by these arrays are illustrated. The possible location of distant centralized noise sources is discussed briefly.

The Grounded Condenser Antenna Radiation Formula.[19] W. HOWARD WISE. Exact formulas for the wave function and vertical electric field at the surface of the ground are derived for a vertical dipole of zero height.

[18] *Proc. I. R. E.*, October, 1931.
[19] *Proc. I. R. E.*, September, 1931.

Contributors to this Issue

O. B. BLACKWELL, B.S. in Electrical Engineering, Massachusetts Institute of Technology. After graduation, he entered the Engineering Department of the American Telephone and Telegraph Company and in 1919 was made Transmission Development Engineer. Mr. Blackwell has general supervision of transmission developments and has been prominently associated with progress in long distance wire and radio telephony.

KARL K. DARROW, B.S., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., University of Chicago, 1917; Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Dr. Darrow has been engaged largely in writing on various fields of physics and the allied sciences. Some of his earlier articles on Contemporary Physics form the nucleus of a book entitled "Introduction to Contemporary Physics."

CLIFFORD E. FAY, B.S. in Electrical Engineering, Washington University, 1925; M.S., 1927. Bell Telephone Laboratories, 1927–. Mr. Fay has been engaged principally in the development of power vacuum tubes for radio purposes.

J. B. JOHNSON, B.S., University of North Dakota, 1913; M.S., University of North Dakota, 1914; Ph.D., Yale, 1917; Engineering Department, Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. As Research Physicist, Dr. Johnson has been engaged in the development and application of special vacuum tubes and gaseous discharge devices.

FREDERICK B. LLEWELLYN, M.E., Stevens Institute of Technology, 1922; Ph.D., Columbia University, 1928. Western Electric Company, 1923–25; Bell Telephone Laboratories, 1925–. Dr. Llewellyn has been engaged in the investigation of special problems connected with radio.

D. A. MCLEAN, B.S. in Chemical Engineering, University of Colorado, 1929. Bell Telephone Laboratories, 1929–. Mr. McLean's work in the Chemical Research Department has dealt principally with plasticity and viscosity studies of various materials.

H. Nyquist, B.S. in Electrical Engineering, North Dakota, 1914; M.S., North Dakota, 1915; Ph.D., Yale, 1917. Engineering Department, American Telephone and Telegraph Company, 1917–19; Department of Development and Research, 1919–. Dr. Nyquist has been engaged in transmission work particularly relating to toll cables.

R. L. Peek, Jr., Columbia University, A.B., 1921; Met. E., 1923. Western Electric Company, Engineering Department, 1924–25; Bell Telephone Laboratories, 1925–. Mr. Peek is in the Materials Testing Division of the Apparatus Development Department.

E. E. Schumacher, B.S., University of Michigan; Research Assistant in Chemistry, 1916–18. Engineering Department, Western Electric Company, 1918–25; Bell Telephone Laboratories, 1925–. Mr. Schumacher's work since coming with the Bell System has related largely to research studies on metals and alloys.

The Bell System Technical Journal

April, 1932

A New Key West—Havana Carrier Telephone Cable *

H. A. AFFEL, W. S. GORTON and R. W. CHESNUT

A new submarine cable has recently been laid between Key West and Havana in order to furnish more telephone facilities between the United States and Cuba. The cable has a single central conductor with concentric tape return and employs the newly developed material paragutta for insulation. A carrier telephone system provides three telephone channels. As ultimately developed a still greater number of facilities may be made available over the cable.

IN January 1931, telephone service was inaugurated over a new submarine cable to Cuba, spanning the hundred-mile or more stretch of deep water between Key West and Havana. Telephone service to Cuba dates back about 10 years earlier when three continuously loaded cables were laid between Key West and Havana. These cables and their associated terminal apparatus were completely described in a paper [1] by Martin, Anderegg, and Kendall, presented to the Institute at its midwinter convention in 1922. The new cable, like the earlier ones, is owned by the Cuban-American Telephone and Telegraph Company, an organization controlled jointly by the American Telephone and Telegraph Company and the Cuban Telephone Company for the purpose of providing telephone facilities between the United States and Cuba.

In the past decade the communication art has advanced in many respects so that three telephone circuits are provided by carrier operation, using high frequencies, over this single improved type cable which is not much larger than one of the three earlier cables.

The three telephone channels now made available are connected to carrier telephone channels operating on open-wire lines northward to Washington and thence over four-wire cable circuits to New York where they terminate as New York-Havana circuits. The telephone circuits derived from the three older cables now terminate at Miami or Key West, where they may be switched to other distant points.

The new cable was designed to satisfy economically the initial need for more circuits by an adaptation of standard carrier apparatus. The design was made sufficiently liberal so that when still more facilities are required certain further development work should make it possible to obtain them over the same cable. From a transmission standpoint, the feature of most interest is the unusually low receiving levels at which operation is successfully carried on.

---

* Presented at A. I. E. E. Midwinter Convention, Jan. 25–29, 1932, New York, N. Y.

[1] "Key West—Havana Submarine Telephone Cable System," *A. I. E. E. Trans.*, Vol. 41, p. 1, 1922.

The Cable

The new cable, which has been designated the 1930 cable, is the longest deep sea telephone cable in existence and is also unique in being the longest telephone cable circuit without intermediate repeaters and without inductive loading. It is somewhat longer (3.7 nautical miles *) than the longest of the 1921 cables and is operated over a far wider frequency range. The new cable operates at frequencies up to about 28,000 cycles per second, and can operate up to a still higher frequency, whereas the old cables are operated only up to 3,800 cycles per second. The longest deep sea carrier frequency cable before the laying of the present cable was that connecting Tenerife with Gran Canaria in the Canary Islands. This cable [2] is non-loaded and is intended to utilize approximately the frequency range now utilized by the new cable but is much shorter, being only 39.7 nautical miles in length.

Paragutta

The feature of the new cable which has enabled this great improvement to be attained is the insulation, which is of paragutta. This material was developed at the Bell Telephone Laboratories and is composed of deproteinized rubber, deresinated balata, and wax. It has been described in detail by A. R. Kemp.[3] Heretofore submarine cables having waterproof insulation, with the exception of the Catalina Island cables [4] which are insulated with a special rubber mixture, have almost invariably been insulated with gutta percha, or balata, or a mixture of these substances. Paragutta has better electrical properties than any of these materials.

Some idea of the improvement represented by paragutta can be had from Table I which lists the significant a-c. electrical properties of representative submarine cable insulations. It is evident from these figures that the use of paragutta effects a considerable reduction in the size of a cable for a given attenuation, not only for telephone cables but also for telegraph cables. This decrease in size is due both to the smaller dielectric constant and to the smaller leakance.† The smaller dielectric constant is equally effective in reducing the size of both loaded and non-loaded cables. The smaller leakance is especially important in the case of loaded cables. The decrease in size due to the use of paragutta varies, of course, with the size of the cable. Some idea of its amount may be obtained from the fact that a cable insulated with gutta percha of the sort used in the 1921 cables would weigh 45 per cent more and cost about 65 per cent more than the new cable,

---

* One nautical mile = 6087 feet (1855 meters).
= 1.1528 statute miles.

[2] "Tenerife-Gran Canaria and Algeciras-Ceuta Submarine Cables," K. E. Latimer and J. R. Vezey, *Electrical Communication*, Vol. 9, p. 226, 1931.

[3] "Paragutta, a New Insulating Material for Submarine Cables," *Journal of the Franklin Institute*, Vol. 211, p. 37, 1931.

[4] "Carrier Current Communication on Submarine Cables," H. W. Hitchcock, *A. I. E. E. Transactions*, Vol. 45, p. 1169, 1926.

TABLE I

COMPARATIVE ELECTRICAL PROPERTIES OF GUTTA PERCHA AND PARAGUTTA INSULATIONS OF SUBMARINE CABLES AT 22 KILOCYCLES PER SECOND UNDER SEA BOTTOM CONDITIONS

	Dielectric Constant	Ratio of Leakance to Capacitance
Gutta Percha (telegraph cable)	3.3	4040
Gutta Percha as used in 1921 Key West-Havana Cables	3.1	—
Gutta Percha as used in Tenerife-Gran Canaria and Algeciras-Ceuta Telephone Cables *	2.92	3815
Paragutta (Key West-Havana 1930 Cable)	2.67	229

* Electrical Communication, Vol. 9, p. 217, 1931.

The use of this new material in the manufacture of a cable gave rise to numerous problems, one of which deserves particular mention. namely that of jointing the paragutta. A new technique of jointing was developed which not only produces good joints in paragutta-insulated cable but also produces better joints in gutta percha-insulated cable than can be made by the conventional process which has been in general use since gutta percha cables were first manufactured.

Cable Design

The 1930 cable is similar in type to the 1921 cables except that it is not loaded. It is provided with copper return tapes [5] and also with a thin copper tape under the return tapes for protection against marine organisms. The principle of the electrical design of the cable was to strike an economic balance between all of the factors affecting the attenuation and thus to secure a cable of the desired attenuation at the lowest possible cost. The result of this procedure is a core structure having a much greater thickness [6] of insulation compared with the diameter of the central conductor than is the case with low-frequency cables, either telephone or telegraph. The particular factor contributing to this result is the large skin effect, which occurs principally in the central conductor but also appreciably in the return tapes. The ratio of the a-c. to the d-c. resistance increases at a rapid rate if the central conductor and return tapes are made heavy in an attempt to reduce the weight of the insulation. At carrier frequencies

---

† The term leakance, as used here, includes all energy losses in the dielectric.

[5] "Transmission Characteristics of the Submarine Cable," Carson and Gilbert, *Journal of the Franklin Institute*, Vol. 192, p. 705, 1921.

[6] British Patent No. 343093, May 7, 1931.

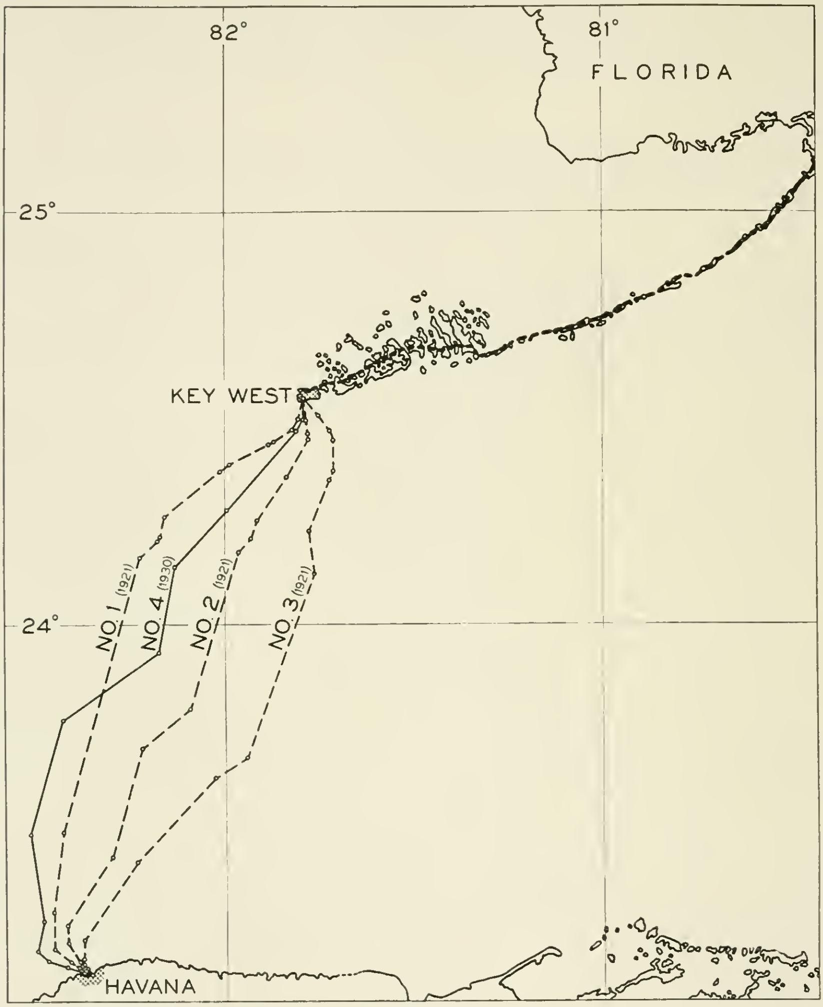


Fig. 1—Submarine telephone cables between Key West and Havana, and U. S. land line connection.

the effect of the armor wire is negligible from the attenuation standpoint. Magnetic modulation due to the presence of the armor wire was investigated carefully and likewise found to be negligible. The lay, or pitch, of the copper return tapes was made much longer [7] than was the case in the 1921 cables. This brought about a substantial decrease in the effective resistance of the tapes themselves and reduced the eddy-current losses which are due to the helical nature of the return tapes.

In all matters of mechanical design accepted cable practice was followed. A discussion of some of the mechanical features of cable design, as well as some account of the general submarine cable problem, has been given by Martin, Anderegg, and Kendall in the paper above referred to.

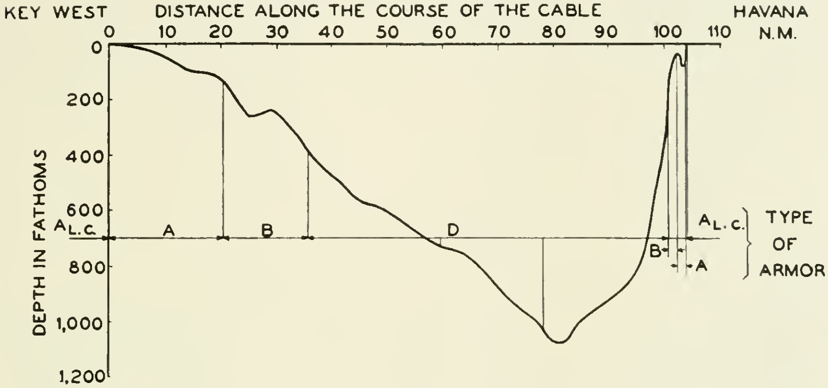


Fig. 2—Depth of water along route of cable.

The new cable is 108.6 nautical miles (125.2 statute miles) in length. It was manufactured by the Norddeutsche Seekabelwerke-A.G. of Nordenham, Germany and laid by their cable steamer "Neptun." Its route is shown in Fig. 1, and the profile of the route in Fig. 2. The maximum depth attained is 1080 fathoms (6480 feet). The particulars of the core structure are given in Table II and those of the armor in Table III. Cross-sections of the various types of cable are shown in Fig. 3. A photograph of three of the types is shown on Fig. 4. The attenuation of the cable and its characteristic impedance are shown in Figs. 5 and 6, respectively. The attenuation is characteristic of a non-loaded cable in that it increases rapidly at low frequencies but less rapidly at high frequencies whereas the reverse is true, in a general way, of the loaded cable. This is shown by the attenuation curve of one of the 1921 cables which is given in Fig. 5 for comparison. The d-c. properties of the laid cable are shown in Table IV.

---

[7] U. S. Patent No. 1700476, January 29, 1929.

TABLE II

Core Structure of Key West-Havana Cable No. 4

Central Conductor	
Central Wire	Diameter .138 in.
Surrounds	6 tapes .0142 in. × .079 in.
Weight of Whole	505 lbs./n.m.
Diameter of Whole	.167 in.
Impregnating Compound	.0024 in. thick
Insulation	
Weight (Including Compound)	677 lbs./n.m.
Outer Diameter	.614 in.
Fabric Tape	.010 in. thick
Protective Tape (Copper)	1 in. × .004 in. with 10% overlap
Weight	209 lbs./n.m.
Return Conductor	6 tapes .319 in. × .019 in.
Weight	845 lbs./n.m.
Outer Diameter	.681 in.
Ozokerite Tape	
Outer Diameter	.704 in.

TABLE III

Armor for Key West-Havana Cable No. 4

Type of Armor	Length Laid n.m.	Armor Wires		Weight of Completed Cable		Outer Diameter of Cable (Inches)
		No. of Wires	Diameter (Inches)	(1) Wet (lbs. per n.m.)	(2) In Water (lbs. per n.m.)	
$A_{l.c.}$	.40	13	.300	29300	21660	1.93
A	21.94	12	.300	24000	17380	1.81
B	17.50	16	.200	16140	11000	1.57
D	68.79	22	.104 e.w.t.	8990	5115	1.30

l.c. = lead covered.
e.w.t. = each wire taped.

TABLE IV

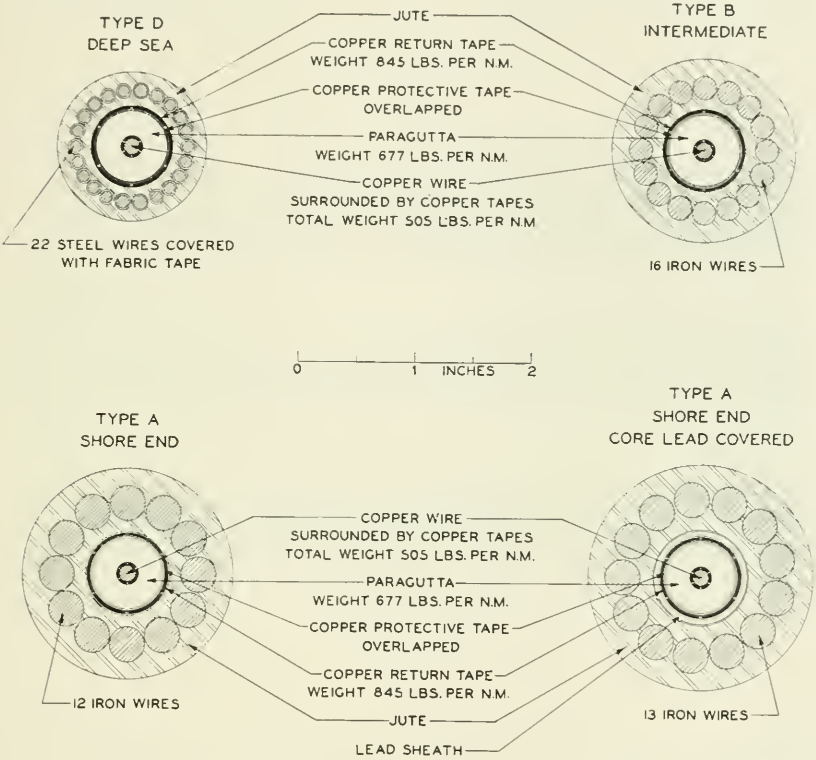
Conductor Resistance, Dielectric Resistance and D-C. Capacity of Key West-Havana Cable No. 4 as Laid

Conductor (Central) Resistance	242 ohms
Dielectric Resistance	1090 megohms
Capacity	22.52 microfarads

One respect in which the present cable differs from the 1921 cables is that it has a single-core termination instead of an unbalanced-type twin-core termination.[8] Experience with the 1921 cables has shown that there is no advantage from the standpoint of the reduction of atmospheric and other extraneous interference to be had from the use of an unbalanced twin-core termination in a cable provided with copper return tapes. A balanced twin-core termination would be effective in reducing interference but certain disadvantages would be connected with its use. Very little shielding beyond that naturally furnished by the sea water is needed and that additional amount is furnished by a wrought iron pipe enclosing the cable between the cable hut and the level of the lowest tides.

---

[8] "Extraneous Interference on Submarine Telegraph Cables," J. J. Gilbert, *Bell System Technical Journal*, Vol. 5, p. 404, 1926.



NOTE:-THE WEIGHTS GIVEN ARE IN POUNDS PER NAUTICAL MILE (1.153 STATUTE MILES).

Fig. 3—Cross-sections of the various types of cable.

Carrier System

As noted previously, in providing carrier apparatus to make use of the high-frequency transmission properties of the new cable, an effort was made to use existing standard types of equipment as far as possible. The three telephone channels are now obtained by an adaptation of carrier apparatus ordinarily [9] used for long-distance transmission over open-wire lines.

The six frequency bands (one for each direction for the three channels) are allocated as shown in Fig. 7. A d-c. telegraph channel is also indicated. This figure likewise shows the frequency allocation of the one telephone and four telegraph channels (three carrier and one direct current) now carried over each of the three older cables. It will be noted that the band width of the new telephone channels is considerably greater than that of the old ones, thus furnishing higher quality speech transmission. In addition, a considerable range of frequencies remains unused on the new cable. This range may be developed when additional message telephone, broadcasting, or telegraph facilities are needed.

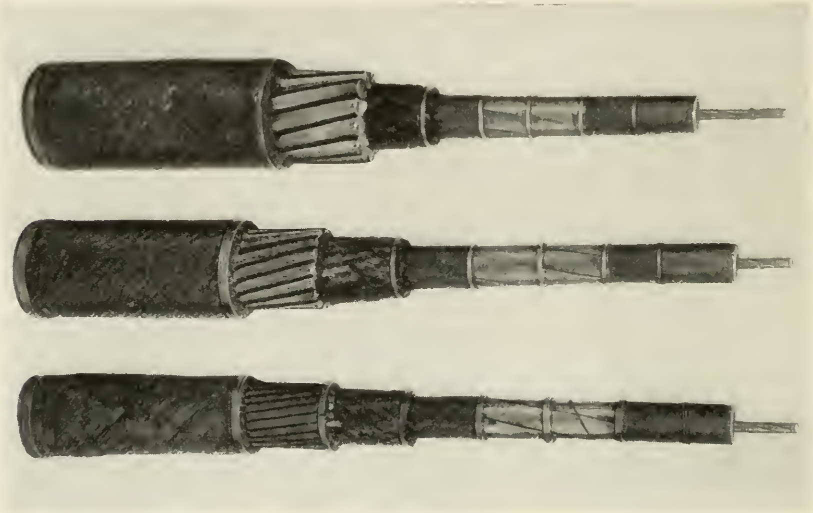


Fig. 4—Shore end (core not lead covered) intermediate, and deep sea types of cable.

In adapting the existing type carrier apparatus for operation over the new cable, the problems consisted chiefly in (*a*) providing for satisfactory transmission over a circuit of considerably higher attenuation than the apparatus was originally designed for and (*b*) providing the necessary transformers to connect together the parts of the circuit having different impedances.

---

[9] "Carrier Systems on Long Distance Telephone Lines," Affel, Demarest and Green, *A. I. E. E. Transactions*, Vol. 47, p. 1360, 1928.

The carrier apparatus is installed in existing telephone offices at Key West and Havana. In each case the office is somewhat over one mile from the cable hut at the water's edge. The submarine cable circuit is connected to the apparatus in the offices through pairs of wires in an underground cable of the paper-insulated lead-covered type, which also carries the circuits of the older cables. In Fig. 8 are shown schematically the connections of the whole cable communication system. The arrangements are practically identical at the two terminals except for differences incident to the fact that different frequency bands are transmitted in the two directions.

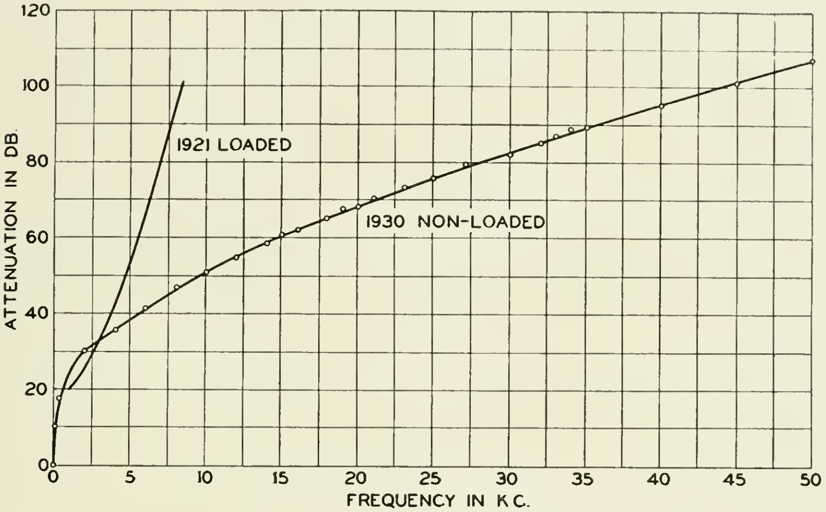


Fig. 5—Attenuation of 1921 and 1930 cables.

Certain coils and condensers are located at the cable hut. A transformer connects the unbalanced 50-ohm submarine cable to the balanced-to-ground 130-ohm pair in the lead-covered cable. The other coils and condensers form a "composite set" which connects the submarine cable to a second pair in the lead-covered cable in order to transmit direct current. This may be used as a d-c. telegraph channel or as an insulation testing circuit.

The carrier equipment in the telephone office may be considered in two categories: (*a*) that which is derived from standard open-wire carrier systems as described in the paper previously referred to, and (*b*) that which is additional and special for this installation.

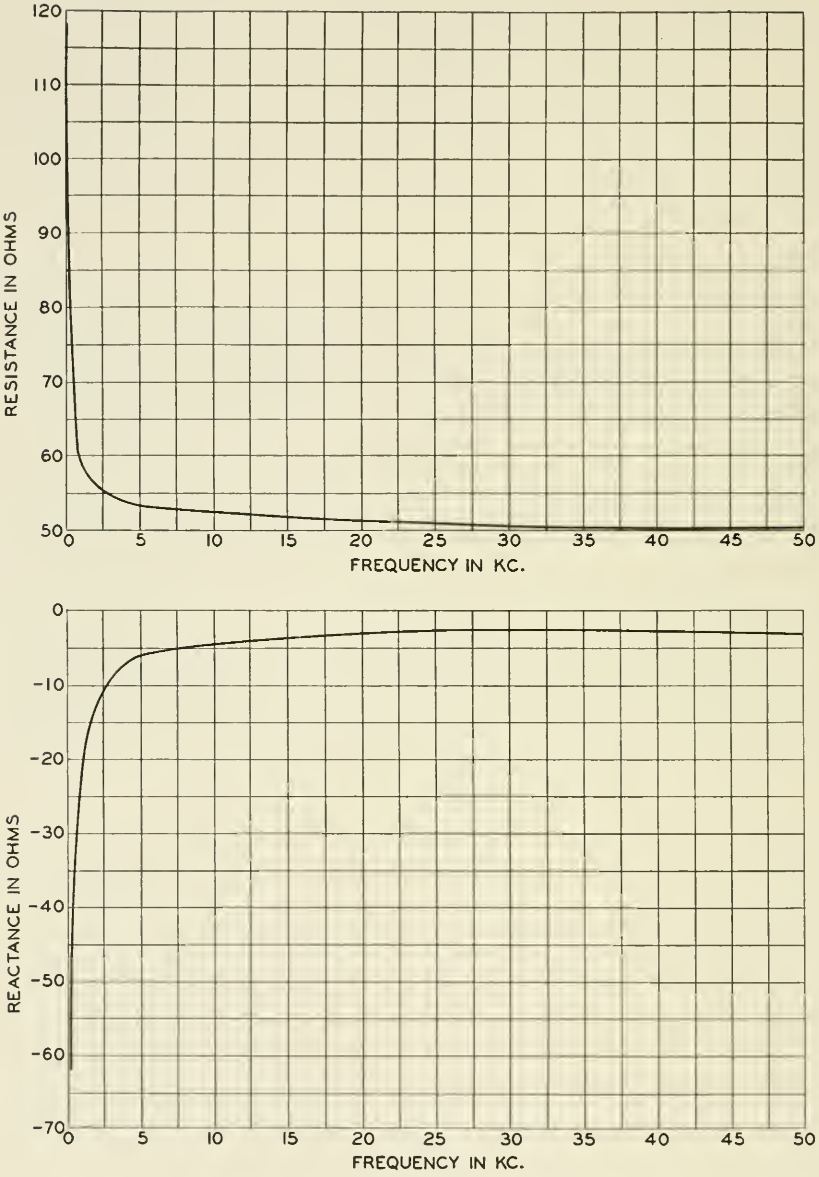
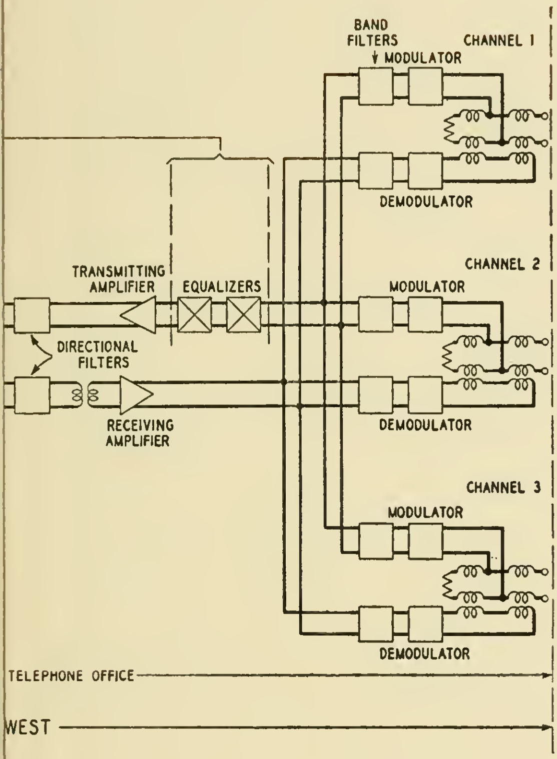


Fig. 6—Impedance of 1930 cable, terminated in its characteristic impedance at Havana, as measured at Key West.


BAND FILTERS
MODULATOR
CHANNEL 1
DEMODULATOR
TRANSMITTING AMPLIFIER
EQUALIZERS
CHANNEL 2
MODULATOR
DIRECTIONAL FILTERS
RECEIVING AMPLIFIER
DEMODULATOR
CHANNEL 3
MODULATOR
DEMODULATOR
TELEPHONE OFFICE
WEST

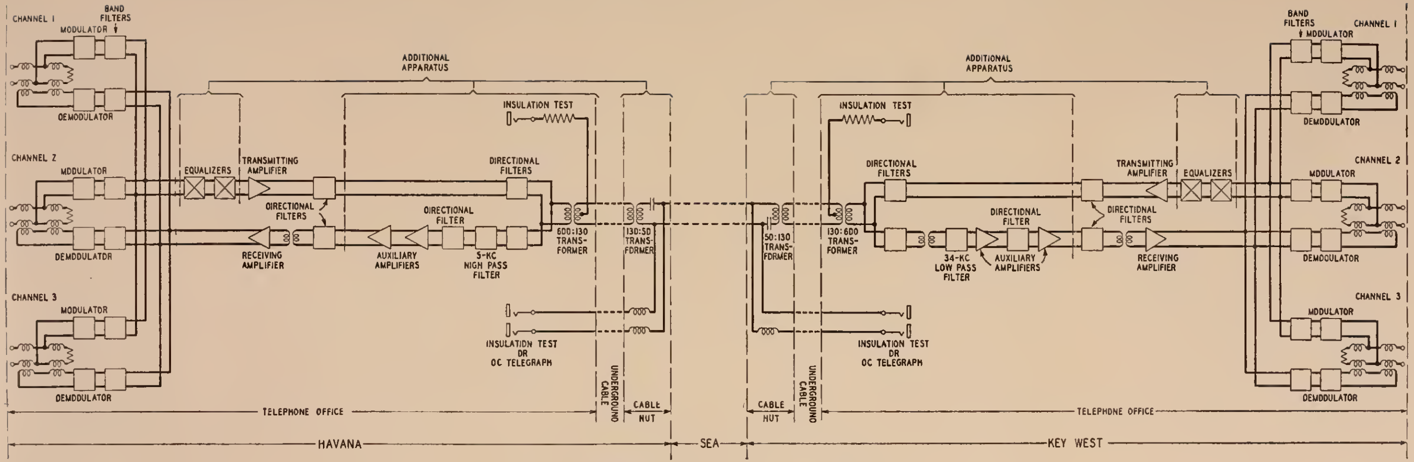


Fig. 8—Schematic circuit of carrier system on 1930 cable.

The additional equipment consists chiefly of added receiving amplifiers and directional filters to care for the greater gain and selectivity needed to operate at the higher attenuation. It includes also equalizers which correct for the varying attenuation of the cable with frequency and a transformer to connect the 600-ohm apparatus to the 130-ohm pair in the underground lead-covered cable. Also, at Key West, a 34-kc. low-pass filter was added to suppress interference from a local radio station having a frequency of about 100 kc. At Havana, a 5-kc. high-pass filter was added to suppress certain relatively low-frequency noises picked up by the circuits in the underground cable.

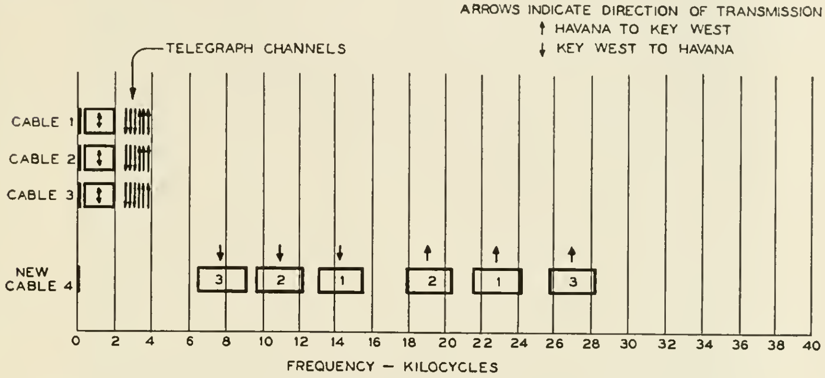


Fig. 7—Frequency allocation of communication channels on Key West-Havana cables.

A photograph of the apparatus installed in the Key West office is shown in Fig. 9. It will be noted that this consists of four "bays" of apparatus, the three nearer of which are practically the same as the carrier equipment ordinarily supplied on long-distance telephone circuits. The fourth bay has the special amplifying and equalizing equipment previously mentioned. A rear view, Fig. 10, shows the interior of one of the special amplifiers, including certain of the special impedance-correcting transformers. Apparatus of a similar nature is installed at Havana.

Certain Transmission Problems

Fig. 11 shows the relative energy of the carrier-frequency speech currents as they traverse the circuit from Havana to Key West. In this direction the higher carrier frequencies are employed. Starting from the left of the diagram, the toll switchboard point, which is taken as zero level, the current of a channel suffers a loss of 9 db in a re-

The additional equipment consists chiefly of added receiving amplifiers and directional filters to care for the greater gain and selectivity needed to operate at the higher attenuation. It includes also equalizers which correct for the varying attenuation of the cable with frequency and a transformer to connect the 600-ohm apparatus to the 130-ohm pair in the underground lead-covered cable. Also, at Key West, a 34-kc. low-pass filter was added to suppress interference from a local radio station having a frequency of about 100 kc. At Havana, a 5-kc. high-pass filter was added to suppress certain relatively low-frequency noises picked up by the circuits in the underground cable.

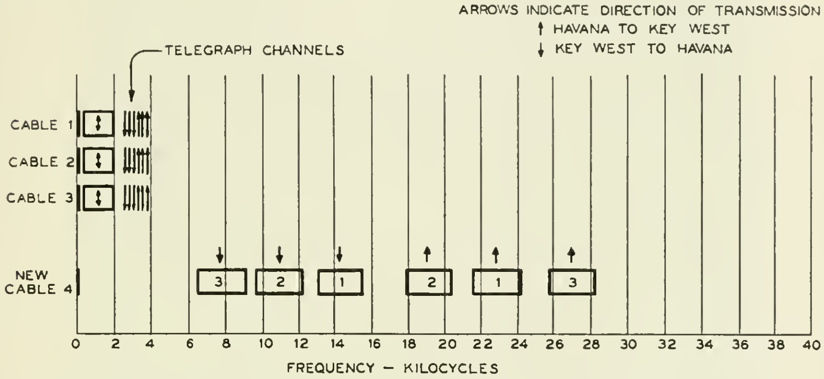


Fig. 7—Frequency allocation of communication channels on Key West-Havana cables.

A photograph of the apparatus installed in the Key West office is shown in Fig. 9. It will be noted that this consists of four "bays" of apparatus, the three nearer of which are practically the same as the carrier equipment ordinarily supplied on long-distance telephone circuits. The fourth bay has the special amplifying and equalizing equipment previously mentioned. A rear view, Fig. 10, shows the interior of one of the special amplifiers, including certain of the special impedance-correcting transformers. Apparatus of a similar nature is installed at Havana.

Certain Transmission Problems

Fig. 11 shows the relative energy of the carrier-frequency speech currents as they traverse the circuit from Havana to Key West. In this direction the higher carrier frequencies are employed. Starting from the left of the diagram, the toll switchboard point, which is taken as zero level, the current of a channel suffers a loss of 9 db in a re-

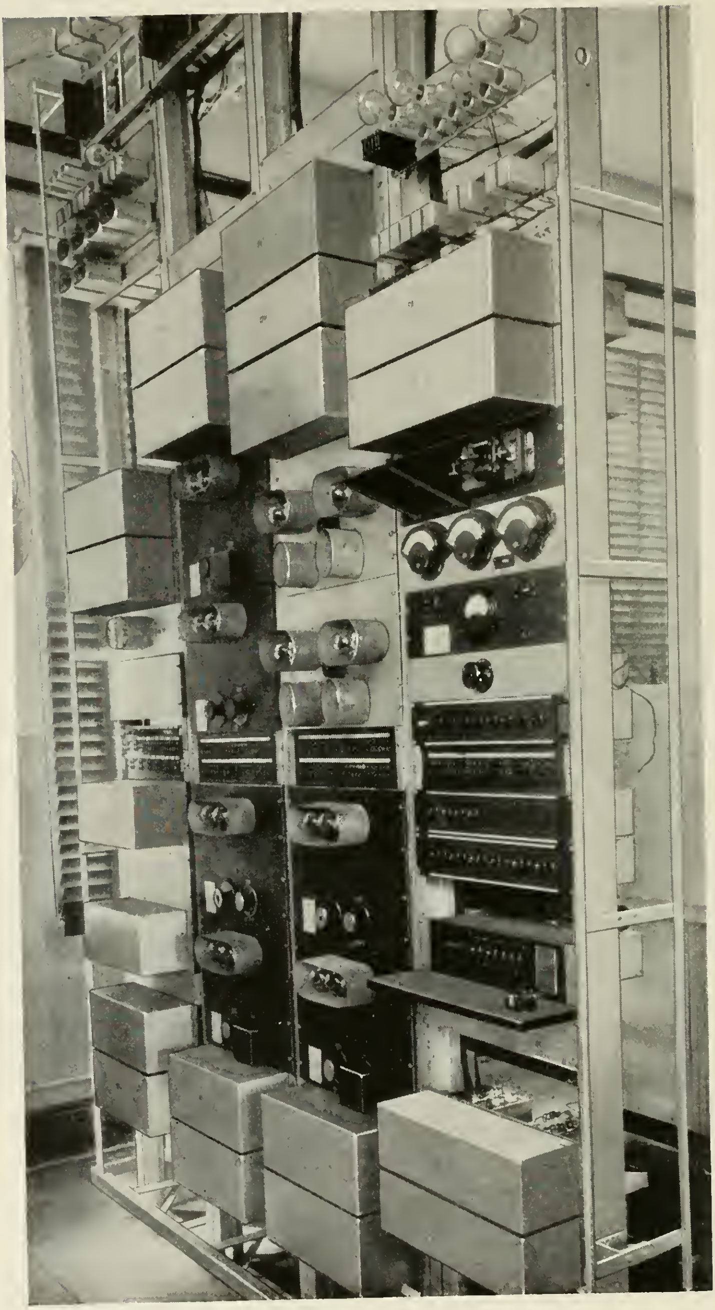


Fig. 9—Carrier terminal apparatus at Key West.

sistance attenuator which is introduced in the circuit to give flexibility in control. This loss can be reduced in value to improve transmission in the case of connections to distant points. As the speech currents enter the carrier apparatus they receive, in addition to the frequency change, an amplification of from 5 to 10 db in the modulator unit and from 10 to 20 db in the transmitting amplifier. The highest frequency channel receives the highest gain and leaves the common transmitting amplifier at a level about 20 db higher than at the toll switchboard.

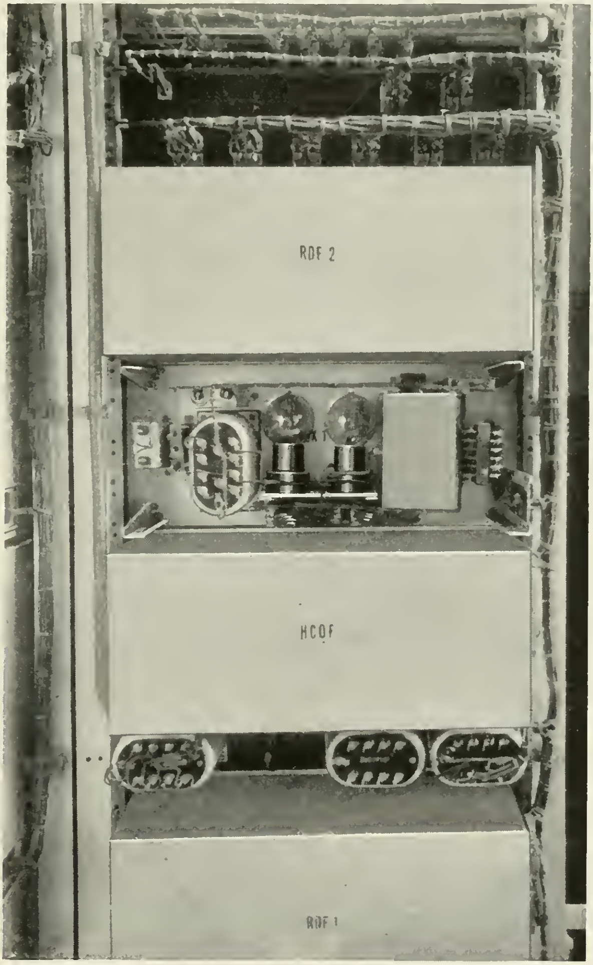


Fig. 10—Rear view of carrier terminal apparatus at Key West showing special amplifier and transformers.

From here the currents leave for Key West by way of the underground cable, the cable hut, and the submarine cable itself.

The highest frequency channel is attenuated somewhat over 80 db by the time it reaches the receiving apparatus at Key West. Here it is stepped up about 70 db by the amplifiers, from the output of which it suffers a few db loss in the demodulator circuit and 9 db loss in the receiving attenuator circuit. If the connection is continued to another toll office, such as Miami or New York, this attenuator is adjusted to give a 9 db loss between the hybrid coil at Key West and the receiving toll switchboard in question.

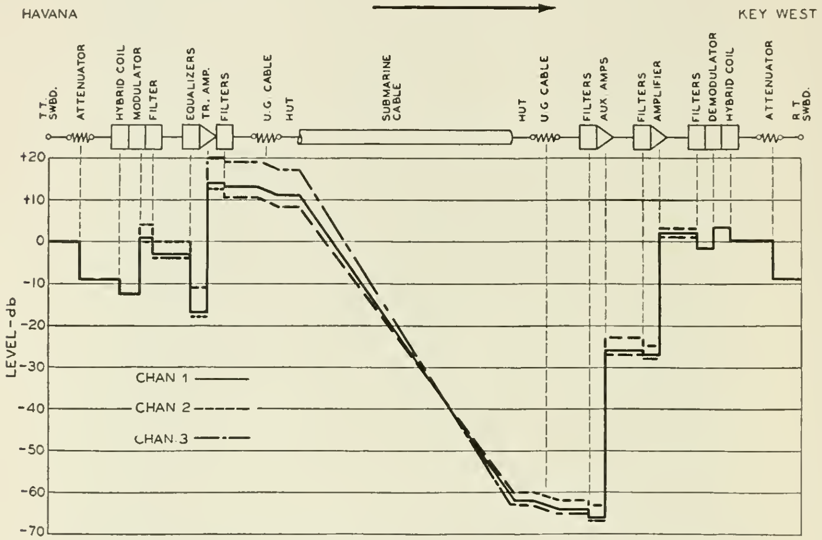


Fig. 11—Relative transmission levels in circuit from Havana to Key West.

The problem of satisfactory transmission of the speech currents in this path involves several specially critical points. The transmitting amplifier must amplify to the high levels required without modulating sufficiently to produce troublesome new frequencies falling within its own group of frequency bands. New frequencies produced by the amplifier and falling outside this group are suppressed by the directional filters and cause no trouble. Additional possible sources of modulation are the directional filters and impedance matching transformers. They must transmit the high level currents coming from the amplifier but must not modulate them sufficiently to produce troublesome new frequencies falling within the oppositely directed group of bands. The latter are at an exceedingly low level and so the modulation in these circuit elements must be kept extremely small. Some ordinary resistances, mica condensers, etc., have been found to have enough modulation to be serious. Ordinary iron core transformers are likely to be very bad. Special design was required to reduce the modulation in the impedance matching transformers to tolerable limits.

Noise Prevention

As noted previously, at the receiving end the important problem is naturally to keep the low level receiving circuit free from interference. In taking the necessary precautions along these lines, a large number of sources of noise were investigated. These included crosstalk from other carrier telephone or telegraph systems, high-frequency oscillations set up by d-c. telegraph apparatus in the vicinity, radio stations, power wires, submarine telegraph cables, and many minor sources.

Space requirements prevent a complete discussion of this work but a tabulation of a few of the remedial measures may be interesting.

Special filters in all telephone office power supply sources.
Special shielding arrangements in the construction of the receiving amplifiers.
Series "choke" coils in all the d-c. telegraph circuits which enter the underground cable.
D-c. telegraph apparatus in the office with special "spark-killer" and high-frequency suppression units.
Frequency limiting apparatus in the Commercial Cable Company's submarine telegraph circuit.
Frequency limiting filters in the carrier telegraph circuits of the 1921 cables.
Special shielding to reduce induction from the other carrier telephone equipment in the terminal offices.
Improvement in the balance (to ground) of some of the apparatus previously installed.
Various special grounds in the apparatus and cable circuits both at the telephone office and in the cable hut.

Of particular interest was the case where the Commercial Cable Company's submarine telegraph circuits from Havana to New York, a communication facility of low inherent frequency range, interfered with the carrier channels, having frequencies up to 28 kc., by induction in the underground cable through which both circuits passed. In this case, by the generous cooperation of the Commercial Cable Company, frequency limiting equipment was added to their cable transmitter.

The application of these various measures served to reduce the total noise manifest at Key West and Havana by a factor of from 40 to 60 db so that as now operated the noise from all other sources is of the same order of magnitude as that picked up in the submarine cable itself, which is extremely small at the high frequencies. It approaches that of the Johnson effect [10] or so-called resistance noise, in which the conductor itself acts as a source of voltage fluctuations which are distributed uniformly over the whole frequency spectrum. The Johnson effect presents a definite lower level limit to all communication circuits.

Future Possibilities

The present arrangements do not represent the ultimate possibilities in communication facilities which the new cable affords. As previously noted, if traffic requirements continue to grow, so that more facilities are required, a wider frequency range may be employed and additional channels obtained. Certain further development work will be required and it is believed that at least three more telephone channels can be provided. In addition, if telegraph circuits are required, carrier telegraph systems may be operated in place of one or more of the telephone channels. For example, if the ultimate capacity of the cable is a total of six telephone channels, a possible arrangement would be to employ two of these telephone channels for carrier telegraph circuits. The cable may then carry simultaneously a total of four telephone messages and 24 or more two-way telegraph messages.

---

[10] "Thermal Agitation of Electricity in Conductors," J. B. Johnson, *Physical Review*, July 1928, p. 97–109.

Cellulose Acetate Treatment for Textile Insulation—Engineering Development

By E. B. WOOD and D. R. BROBST

The development of a cellulose acetate lacquer treatment for textile insulated wire has made available an improved type of wire for telephone central office use. The desired improvement in electrical characteristics is obtained when the textile fibers are laid down and covered by the cellulose acetate film.

The accompanying graphs show the comparative electrical characteristics at various humidities, of wires insulated with commercial and purified cotton and silk servings, before and after treatment with cellulose acetate lacquer.

Introduction

THE improved standards of transmission required for present-day telephone communication have greatly increased the importance of improved electrical characteristics for telephone central office wire insulation. At the same time the tremendous growth of telephone systems, together with the increase in complexity of central office equipment due to the introduction of dial switching apparatus, has increased the quantity of insulated wires required to such an extent that the use of comparatively cheap materials is a matter of large economic importance. Silk and cotton yarns applied in the form of wrappings or braidings have been the standard materials for telephone central office wire insulation for many years. These materials in proper combinations and supplemented in certain cases by enamel and impregnating waxes provide sufficient dielectric strength to withstand the comparatively low voltages employed to operate telephone apparatus. This type of insulation also fulfills certain controlling mechanical requirements, in that it occupies small space, is not easily damaged by normal handling and can be applied in a large number of color combinations. On the other hand, there are disadvantages attendant upon the use of textile insulation, the most serious of which is the wide variation in insulating properties of such materials under different conditions of atmospheric humidity and temperature, caused mainly by changes in the moisture content of the materials.

The efforts which have been made to improve textile insulation in this respect have had a two-fold objective, namely, to provide at moderate cost a super-quality insulation for use where it is important to have the best electrical characteristics obtainable, and to improve cotton sufficiently to permit its use instead of silk as far as possible for general purposes. The latter has considerable direct economic importance because of the large difference in cost of insulating silk and cotton and the quantities of the materials involved.

Within the past three years, two methods of improving the electrical characteristics of textile insulation have been brought into commercial use. The first is the purification of silk and cotton whereby electrolytic impurities such as sodium and potassium salts inherent in the commercial materials are removed by a simple and inexpensive washing process.[1] The second is the treatment of textile insulated wire with cellulose acetate, which is the subject of discussion in this paper and the contemporary paper "Cellulose Acetate Treatment for Textile Insulation-Development of the Manufacturing Process" by Messrs. C. R. Avery and H. Kress.

Cellulose acetate became of interest in connection with insulation problems several years ago, when it was investigated in the form of artificial silk for use as a substitute for natural silk in wire and cable insulation. At that time, the material was found to possess excellent electrical characteristics and satisfactory stability, but it did not prove to be economically satisfactory as a general substitute for silk because the physical characteristics of the yarn made its use on standard high speed insulating machinery difficult. Application of the material in the form of a lacquer to cotton or silk insulation appeared to offer more promise and has proven advantageous, as will appear from the following discussion. The treatment, as now applied to telephone central office wire insulation, consists in the formation of a coating of the material on the textile insulated conductor, by passing the conductor through an acetone solution of pure cellulose acetate and subsequent evaporation of the solvent. Pure cellulose acetate without the addition of a plasticizer is used, because thus far it has been found more satisfactory than a compounded material as regards the controlling requirements for central office wire insulation, namely good electrical characteristics, slow burning properties and stability. Therefore, this discussion is confined to the characteristics and use of the pure material.

PROPERTIES OF CELLULOSE ACETATE

In the investigation of a material to be used for insulating purposes, it is necessary to examine the processes by which the material is manufactured to determine whether there is anything inherent in these processes which would affect the use of the product. This is particularly pertinent in the case of an insulating material which is to be used for a period of twenty years or more, as in a telephone exchange.

---

[1] "The Predominating Influence of Moisture and Electrolytic Material upon Textiles as Insulators," R. R. Williams and E. J. Murphy, *A. I. E. E. Transactions*, April, 1929. "Purified Textile Insulation for Telephone Central Office Wiring," H. H. Glenn and E. B. Wood, *A. I. E. E. Transactions*, April, 1929.

The manufacture of cellulose acetate is described briefly by the following operations:

In the first or acetylation process, cellulose fibers, usually cotton, are treated with glacial acetic acid and acetic anhydride, together with a catalyst such as sulphuric acid, until the fibers are completely acetylated and pass into solution. The acetate obtained at this stage is brittle, of low tensile strength and insoluble in the commercial solvents. Therefore, the solution is subjected to a hydrolizing process in which water is added and the mixture allowed to stand until hydrolysis has been carried to the point at which the cellulose acetate becomes acetone soluble. The acetyl content is somewhat reduced in this step and serves as an index to the extent of hydrolysis. The solution is then poured into water and the cellulose acetate precipitated, after which it is given a purification treatment until the mass is free from acid and then dried in warm air.

The completed product is a porous, flaky mass, white in color, which, when dissolved in acetone, gives a solution nearly colorless but with a slight amber tinge.

From the above outline of the processes of manufacture, the importance of the acetyl content of the product is obvious. If the acetyl content is too high, the material is not soluble in acetone and if the acetyl content is too low, the hydrolysis has been carried too far and the acetate becomes partly soluble in water. Such an acetate would be unsatisfactory for insulation on account of inferior electrical characteristics under humid atmospheric conditions.

By changes in control of the acetylation and hydrolizing processes various kinds of cellulose acetate may be obtained which, with the same general composition and acetyl content, give different viscosities of solution when dissolved in a solvent. For example, films for experimental purposes have been made from cellulose acetates which vary in viscosity as much as a hundred fold with the same proportions of cellulose acetate and solvent. For lacquer and films, a low viscosity acetate is employed, while for plastics, cellulose acetates of high viscosity are usually specified.

For use as insulation, it is necessary that the acetate be stable throughout the life of a telephone exchange. In other words, it must retain its good electrical characteristics and transparency for a period of twenty years or more though exposed, as it will be, to variations of temperature, indoor sunlight exposure and atmospheric moisture. From the standpoint of stability, especially as regards a possible increased rate of deterioration with time, it is important that the material shall be essentially free from impurities which might be introduced in the acetate manufacturing process.

Cellulose acetate film has very desirable electrical properties characterized by high dielectric strength, low conductivity and low a-c. capacitance and conductance. It absorbs much less moisture than silk, cotton or wool. It has a specific gravity of about 1.25 and a dielectric constant of from 5.5 to 6.0 at 1,000 cps. under atmospheric conditions of 70° F. and 50 per cent relative humidity.

The acetate film is strong and tough and not easily injured by handling. The transparency of the film is such that the colored threads used in the color scheme for identification purposes in telephone wires and cables can be readily seen through the acetate coating.

Cellulose acetate film is very stable under normal conditions and when exposed to artificial aging tests. Tests made in the Laboratories with acetate film exposed to high humidities and high temperatures for several years indicated that there was very little deterioration of the film in its electrical or other physical properties. The electric characteristics of the film were not appreciably affected by this exposure and no discoloration of the film was apparent.

As compared to cellulose nitrate, cellulose acetate is a much more desirable material on account of its slow-burning characteristics, and the fact that the gases given off on the combustion of the acetate are comparatively non-toxic. From these standpoints the hazards involved in the use of the nitrate preclude its use in the telephone central office. Acetate film does not turn yellow with age to the same extent as the nitrate film.

Pure cellulose acetate film is somewhat hard and brittle. This is, of course, a disadvantage because it tends to make the treated wire less flexible than wire with untreated insulation which introduces new problems in the handling of the treated wire. A large amount of work has been done with a view to obtaining a plasticizer for cellulose acetate which will add the property of flexibility to the film without affecting the desirable characteristics which the pure acetate film now possesses. The problem of obtaining such a plasticizer is difficult inasmuch as the general tendency of such materials is to impair the electrical characteristics, lower the tensile strength, and increase the inflammability of the film when used in amounts sufficient to produce a film of desired flexibility.

TREATMENT OF TEXTILE INSULATED WIRE

The cellulose acetate treatment of wire consists essentially in passing the textile insulated conductor through a thin solution of cellulose acetate dissolved in acetone, then through a wiping die to remove the excess lacquer and finally into a heated drying chamber where the solvent is evaporated. This process is repeated several times, usually six, to build up a film of satisfactory thickness and smoothness. The application of heat in the drying process is necessary for two reasons. First, the evaporation of the solvent tends to lower the temperature of the wire considerably and if the temperature falls below the dew point of the surrounding air, moisture will condense on the wet lacquer film, causing it to turn white and opaque. In the second place, the evaporation of the solvent must be rapid in order that the speed of the wire through the lacquering machine may be such as to make application of the treatment to large quantities of wire commercially practicable.

In the earlier stages of the investigation, the insulated wire was thoroughly dried before being treated to eliminate the moisture in the textile. In addition, the wire was treated under vacuum with the object of thoroughly impregnating the whole textile covering and preventing entrance of moisture into the textile after the impregnating process was completed. However, it was found that even with vacuum impregnation, the cellulose acetate did not penetrate the insulation to an appreciable depth, although the solvent appeared to penetrate to the conductor and thoroughly wet the insulating materials. Also, it was found that the coating of cellulose acetate did not prevent the entrance of moisture into the textile to any appreciable extent.

In consideration of these facts, it was concluded and confirmed by tests that the improvement in electrical characteristics of cellulose acetate treated wire under conditions of high humidity is due mainly to the barrier of high resistance lacquer film interposed in the leakage paths formed by moisture in the textile insulation. It may be seen from Fig. 1 that the fibers of untreated cotton insulation project in all directions, and in a twisted pair, interweave to increase the effective area of contact between the conductors and provide a medium for direct leakage paths when moisture is present. With the fibers smoothed down and covered by the application of several layers of lacquer film, the effective area of contact is decreased and any leakage which takes place must be either through or across this relatively high resistance film. In confirmation of this conclusion, it has been found by repeated tests, that a very reliable indication of the improvement in electrical characteristics which may be expected in a treated wire is obtained by observing the extent to which the textile fibers have been laid and covered by the lacquer film. If the surface of the wire is smooth and practically free from projecting fibers as is shown in the photograph, the wire may be expected to exhibit normal improvement.

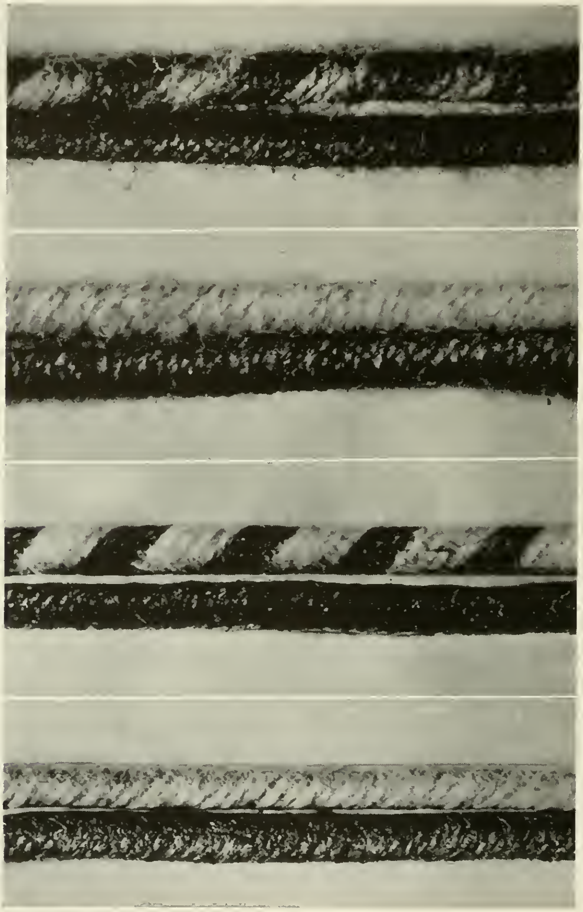


Fig. 1—Textile insulated 22-gauge wires before and after treatment with cellulose acetate. The cellulose acetate covering prevents direct contact between textile fibres of adjacent wires and reduces current leakage.

For example, when textile insulated conductors are twisted and then treated with cellulose acetate, practically no improvement in electrical characteristics of the pair is obtained, because the interlocking fibers of the two conductors which provide the direct leakage paths are not separated. A similar effect has been observed in the treatment of coils wound with textile insulated wire. If the treatment is applied to the wire before winding, the desired improvement in stability of the constants is obtained. Coils treated after winding, however, show very little improvement since the lacquer does not penetrate but merely provides a superficial covering of film which does not exclude moisture or break up the interlocking fibers between turns.

It has been found that increasing the thickness of the lacquer film, beyond that required to cover the fibers and provide a smooth surface, results in relatively small additional improvement in the insulation. This is of economic importance since, with proper methods of application, a relatively thin film may be practically as effective as a thick one requiring a considerably greater quantity of material, and a rapid check on the quality of the product can be made by visual inspection of the surface condition of the treated insulation.

ELECTRICAL CHARACTERISTICS OF TREATED WIRE

The accompanying graphs show a comparison of the electrical characteristics of untreated and treated cotton and silk insulation, respectively, for a cycle of relative humidity ranging from 65 per cent to 90 per cent and back to 65 per cent at a constant temperature of 85° F. The comparison is given for both commercial and purified materials as a matter of general interest, although purified textiles are now used exclusively in Bell System central office wire insulation. The graphs are plotted from data on samples of wire insulated with silk and cotton taken at random from stocks of commercial and purified materials and treated with cellulose acetate under conditions of regular production. The values given by the graphs should not be considered as applying quantitatively to any standard type of central office wire but are intended to show, on a comparative basis, the extent to which the commercial and purified materials have been improved by treatment with cellulose acetate, and the rather remarkable improvement in the characteristics of commercial textiles by both purification and cellulose acetate treatment.

Perhaps the comparison of greatest general interest is that of insulation resistance, Figs. 2 and 3, since it is important in any electric circuit that the insulation shall be capable of preventing undue energy loss from direct current leakage. From these graphs, it is seen that the insulation resistance of commercial cotton may be improved from 100 to 300 fold by treatment with cellulose acetate and in the order of 500 to 2,000 fold by purification plus acetate treatment depending upon the relative humidity. Thus, as indicated by insulation resistance, acetate treated purified cotton becomes a comparatively high

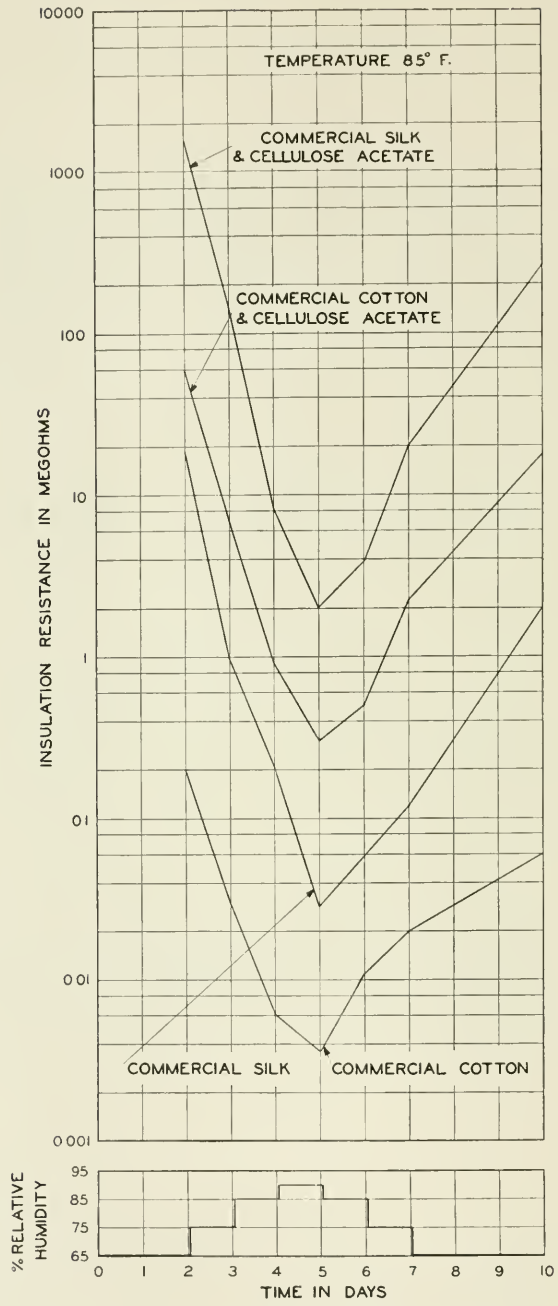


Fig. 2—D-C insulation resistance of 50 feet of twisted pair 22-gauge wire insulated with double servings of equal thickness.

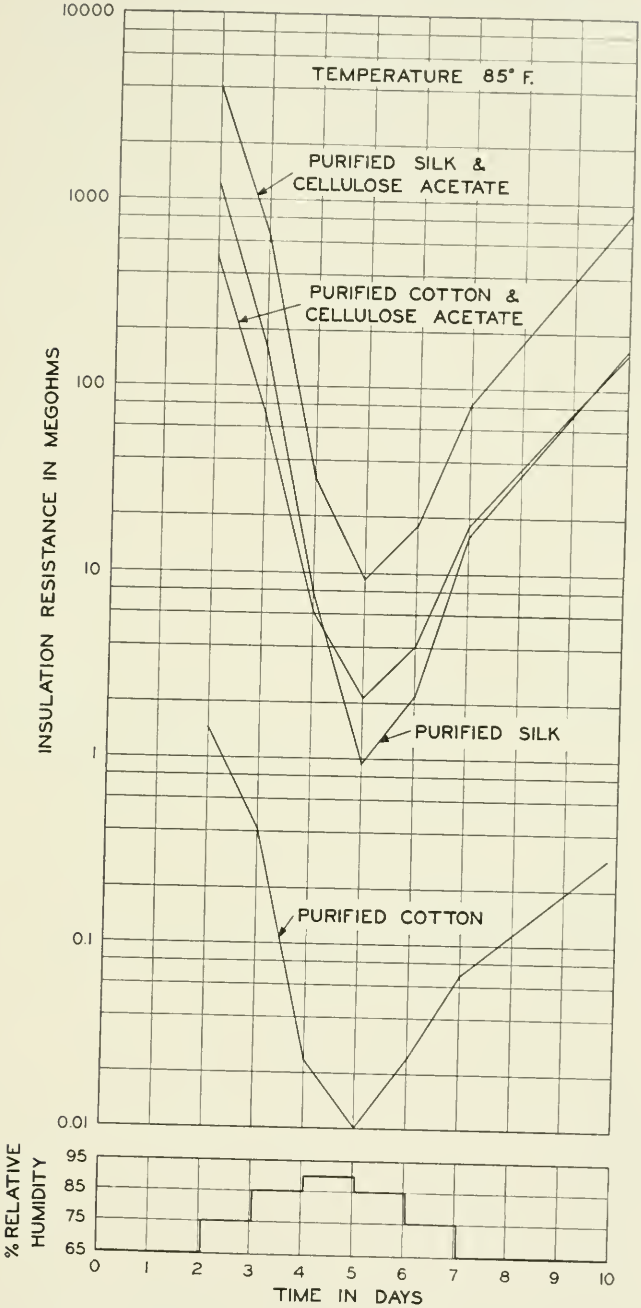


Fig. 3—D-C. insulation resistance of 50 feet of twisted pair 22-gauge wire insulated with double servings of equal thickness.

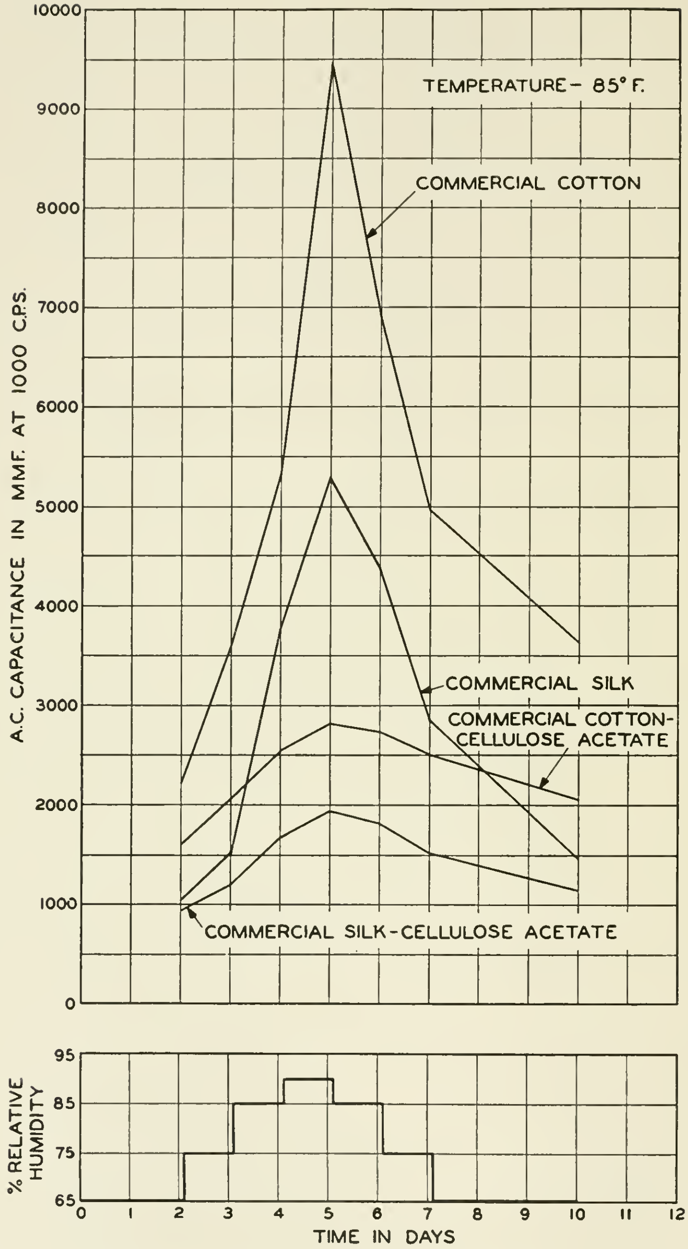


Fig. 4—A-C. capacitance of 50 feet of twisted pair 22-gauge wire insulated with double servings of equal thickness.

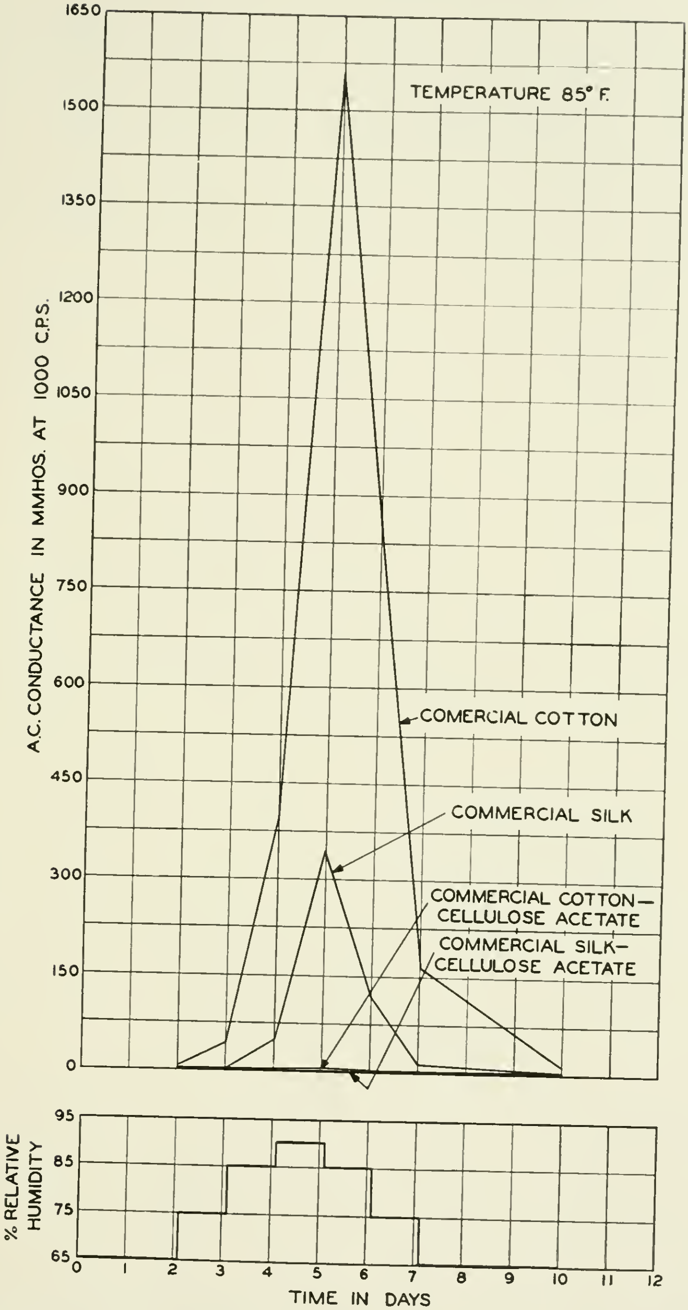


Fig. 5—A-C. conductance of 50 feet of twisted pair 22-gauge wire insulated with double servings of equal thickness.

grade insulation suitable for many purposes where more expensive combinations of silk and cotton have been required heretofore.

From the telephone transmission standpoint, the a-c. characteristics of capacitance and conductance are of particular importance because they determine the loss in transmission of energy at voice and carrier frequencies, which must be kept at a minimum to maintain high quality of telephone communication. A comparison of a-c. capacitance and conductance at a frequency of 1,000 cycles for commercial and purified silk and cotton with and without cellulose acetate treatment is shown in Figs. 4, 5, 6 and 7. The data represented by these graphs converted into transmission loss units are shown in Figs. 8 and 9. These data derive their main significance from the large reduction in capacitance and conductance at the higher humidities, and the fact that it is the variation of these transmission loss characteristics which is of the greatest importance from the telephone transmission standpoint. Losses, if fixed in value, can be compensated for, but if they are subject to wide variations such as those illustrated by the samples of

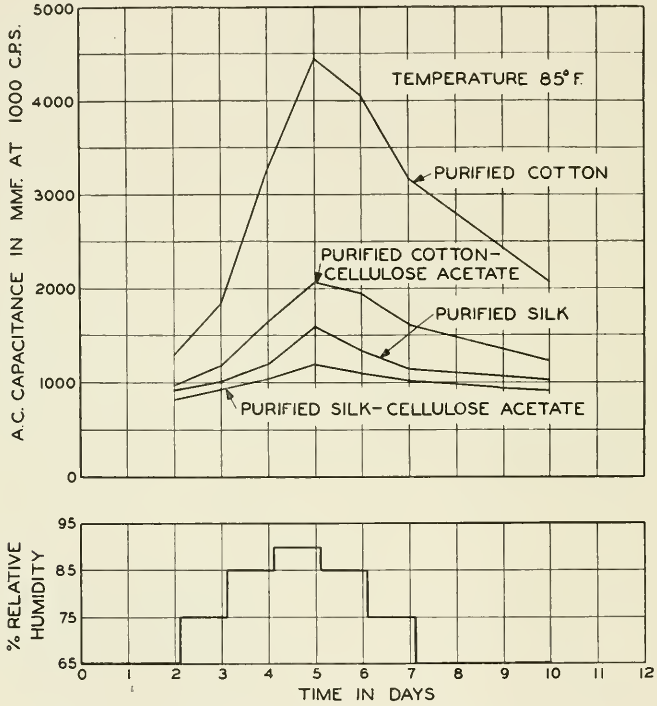


Fig. 6—A-C. capacitance of 50 feet of twisted pair 22-gauge wire insulated with double servings of equal thickness.

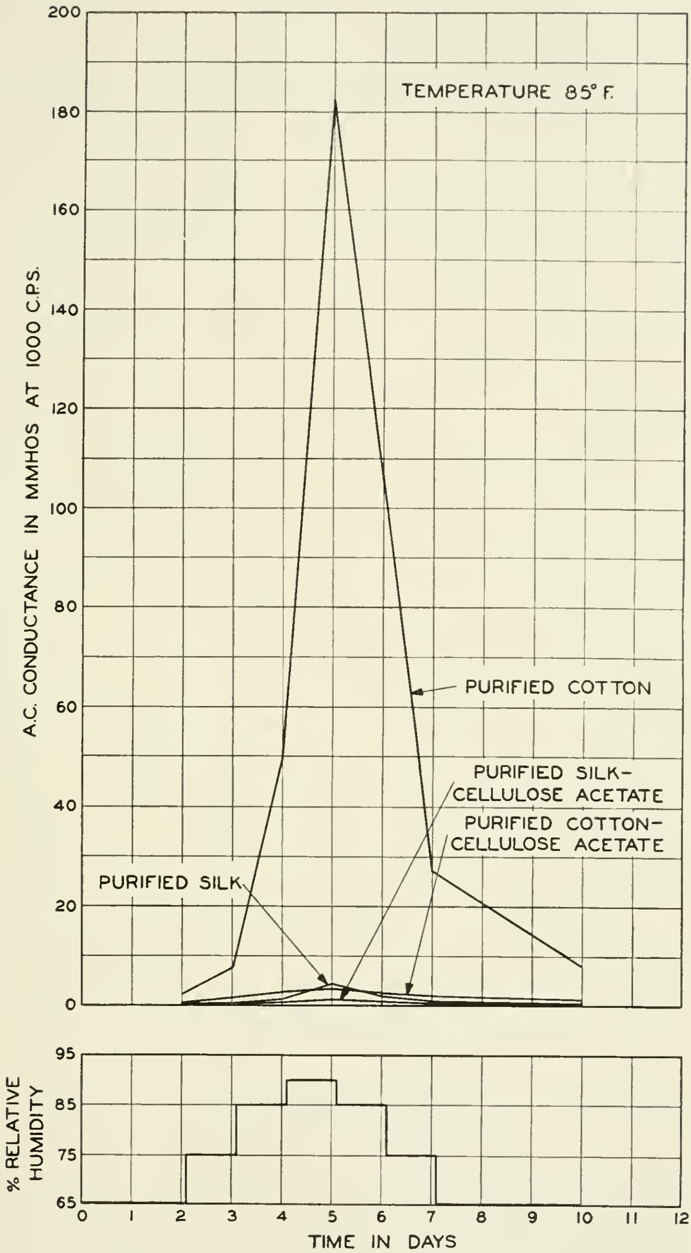


Fig. 7—A-C. conductance of 50 feet of twisted pair 22-gauge wire insulated with double servings of equal thickness.

untreated commercial textiles, the matter of compensation becomes difficult or entirely impracticable.

In addition to maintaining transmission losses at a minimum, it is required in certain toll apparatus that the capacitance and conductance of the parts of the electric circuits be balanced to prevent interference between adjacent circuits which would increase the noise level and impair the quality of voice transmission. Such circuits are usually wired with four-conductor wire and it is required that the characteristics of the insulation and the spacing of the conductors shall be suffi-

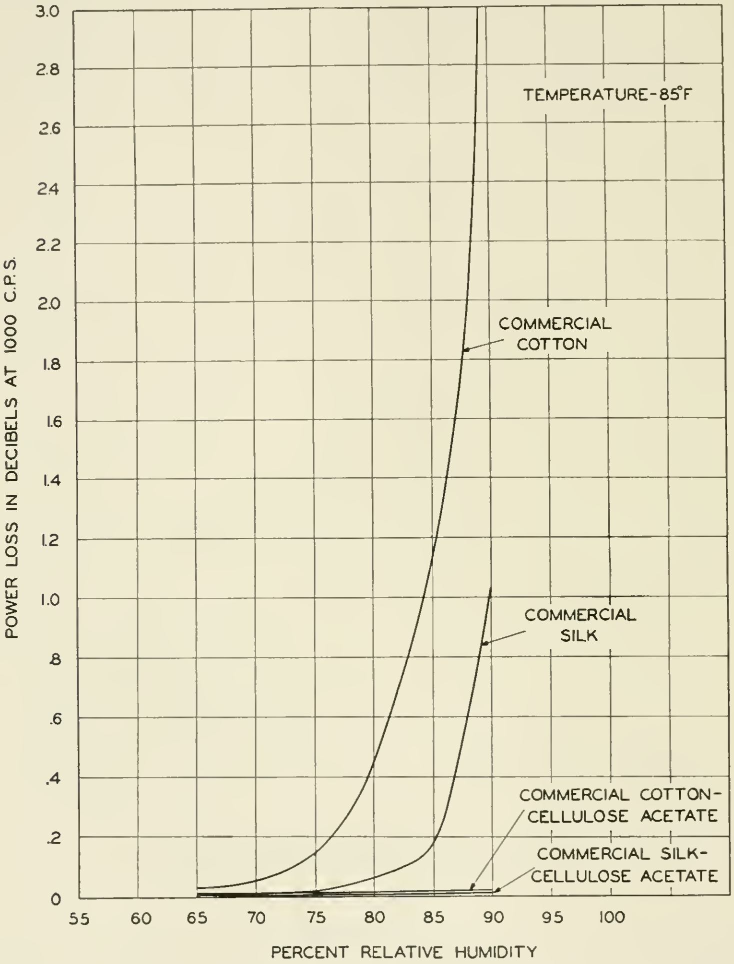


Fig. 8—Transmission loss in 50 feet of twisted pair 22-gauge wire insulated with double servings of equal thickness.

ciently uniform to preserve an electrically balanced circuit. From Figs. 4, 5, 6 and 7, it is seen that the capacitance and conductance of commercial silk and cotton insulation are greatly reduced by cellulose acetate treatment and that the purified materials are also improved considerably in that respect. It is evident, therefore, that the addition of cellulose acetate treatment to purified textile insulation will result in greater uniformity in the electrical characteristics of the product than is possible by use of purified insulation alone, since the effects of any lack of uniformity in the purified material, due to variable results

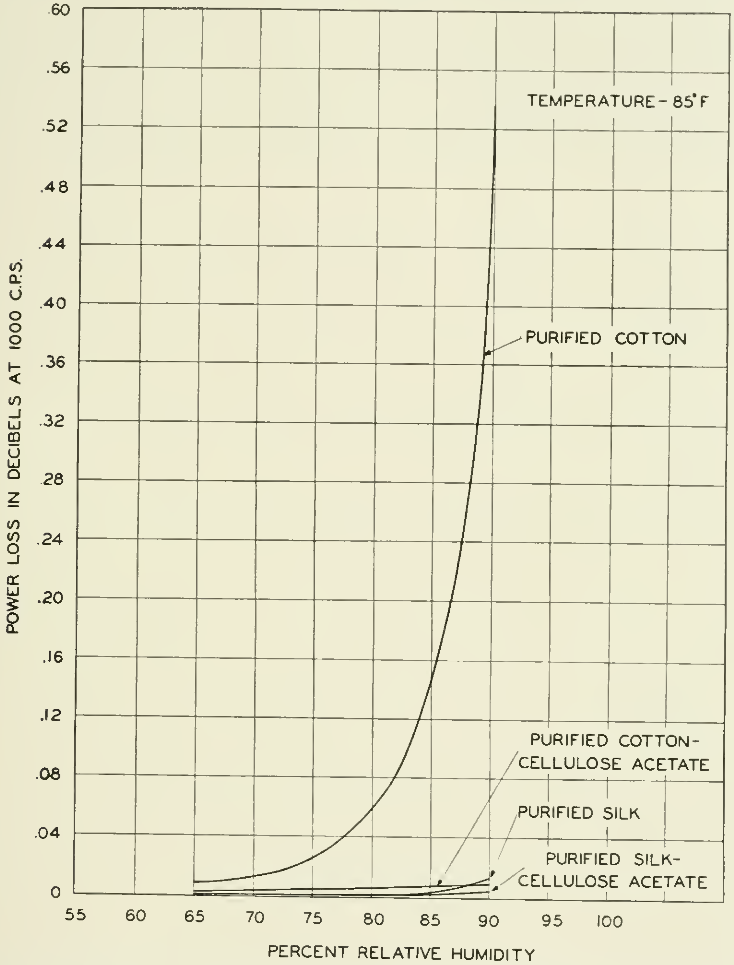


Fig. 9—Transmission loss in 50 feet of twisted pair 22-gauge wire insulated with double servings of equal thickness.

in the purifying process, will be practically nullified by the acetate treatment.

An interesting example of how cellulose acetate treatment improves the electrical balance of a circuit is given in Fig. 10, which shows the capacitance unbalances between phantom and side circuits in a four-conductor wire for toll use. The capacitance unbalance is the main cause of electrical interference between the two circuits mentioned above and it is desirable to have this value as low as possible. Although there are some differences in design in these two types of wire, practically all the improvement is due to the acetate insulation.

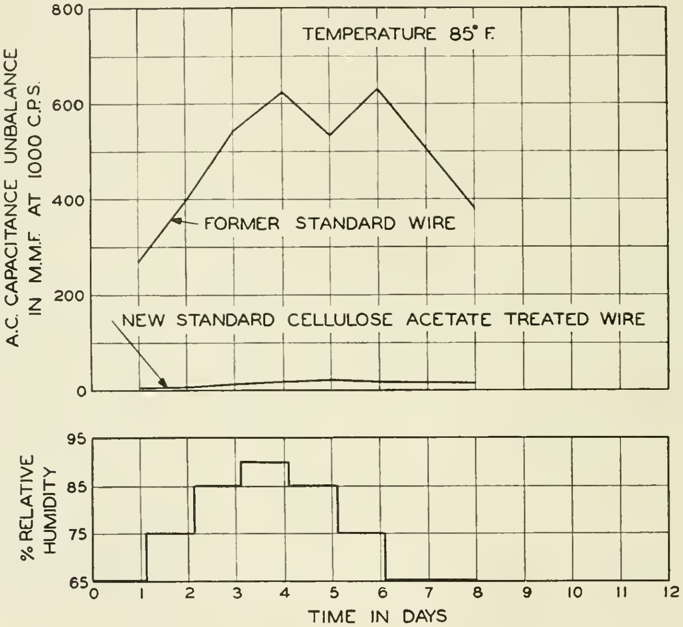


Fig. 10—A-C. capacitance unbalance between the phantom and side circuits of 50 feet of quadded 22-gauge wire.

Application to Apparatus

Advantages in the use of cellulose acetate treated insulation are derived from several sources of which the most important is the improvement in electrical characteristics of cotton and silk, as illustrated by the foregoing graphs. For example, this improvement is sufficient in many cases to permit the substitution of cotton for silk, with a resulting substantial reduction in cost. In other cases, silk has been retained and a cable of much higher quality has been made available for use in toll equipment where the best electrical characteristics obtainable are needed.

The improved electrical characteristics of cellulose acetate treated insulation also make possible the elimination of enamel in a large amount of wire and cable where it has formerly been required to prevent excessive current leakage under conditions of high humidity. This is of economic importance because of the difficulty of removing the enamel preparatory to soldering the wire to terminals, and the precautions necessary to prevent trouble from faulty soldered connections, which increase considerably the cost of installation and maintenance of apparatus.

In distributing frame wire the use of cellulose acetate treated insulation has been found to be particularly advantageous. This type of wiring cannot be installed permanently in cabled form, as is the practice with practically all other types, because of controlling equipment and service conditions, and in order to guard against fire hazard from a large mass of loose wiring, the insulated conductor has, heretofore, been covered with a cotton braid impregnated with flameproofing salts. These salts, because of their hygroscopic and electrolytic nature, have a deleterious effect on the electrical characteristics of the insulation under humid conditions and introduce the danger of excessive leakage and corrosion, particularly near terminals. Exhaustive tests have proved that cellulose acetate treated insulation, without the addition of flameproofing salts, will be as satisfactory as the old type with respect to safety, and the elimination of salts has made it possible to design a wire which is greatly superior to the old type electrically, considerably less expensive to manufacture and smaller in size.

Another advantage in cellulose acetate treatment is its effect in preventing unwrapping and fraying of the textile at terminals. With the old standard wires, fraying is prevented by impregnating the insulation near the ends with wax. This wax treatment is undesirable in that it adds to the flammability of the insulation, tends to obscure the marking colors and collects dust. Cellulose acetate treatment binds the insulation against fraying and provides a smooth glossy surface which does not collect dust readily.

These examples serve to illustrate the more important factors in favor of cellulose acetate treated wire with regard to its application in telephone apparatus. On the other hand, this type of wire has a tendency to be somewhat stiff and springy with the result that its behavior in the operations of twisting, stranding and forming into cables differs considerably from that of the old untreated types. This has made necessary the development of modified manufacturing and installation methods in connection with these operations and, in certain cases, the observance of special precautions. For examples, in the operation of twisting to form pairs, triples and quads, it is necessary to avoid appreciable stretching of the conductors, as this would crack and loosen the acetate film and impair the appearance of the wire. Fortunately, cracking of the film does not affect the electrical characteristics appreciably, so long as it is not severe enough to permit interlinkage of textile fibers, as discussed in a previous paragraph.

The first application of cellulose acetate treated wire on a regular production basis was made early in 1930, although considerable quantities were installed for service trials in commercial apparatus in 1927. In order to take advantage of the improved electrical characteristics where they are of greatest value, cables and wire used in the toll plant are being changed to employ the new insulation first. Supplementing this program, consideration is being given to extending the use of acetate treated insulation to include wire aggregating annual requirements of the order of three billion feet for the local plant. Application of the new type wire is being made gradually in order that manufacturing and installation methods and technique may be further developed as required in connection with this program.

Cellulose Acetate Treatment for Textile Insulation—Development of the Manufacturing Process

By C. R. AVERY and H. KRESS

Equipment was developed and a plant constructed for coating textile insulated wire with a film of cellulose acetate. The wire is treated at a speed of 240 feet per minute. Recovery of more than 85 per cent of the acetone used as a solvent is effected with carbon adsorbers. Thorough precautions have been taken to prevent fire and explosion and to render them harmless if they should occur.

Introduction

IN the telephone plant of the Bell System a large quantity of textile insulated wire is employed for wiring the switchboards and connecting them to the incoming lead covered cables. The insulation on this wire varies with its use in the plant, common constructions in the past being tinned copper wire insulated with two servings of silk and one of cotton, or, where the requirements are not so exacting, enameled wire with two servings of cotton.

The constant demand for better electrical characteristics in telephone circuits has led the engineers of Bell Telephone Laboratories to seek an improved insulation for this wire as discussed in the contemporary paper, "Cellulose Acetate Treatment for Textile Insulation, Engineering Development," by E. B. Wood and D. R. Brobst. The old design wire was affected by the variation in the dielectric properties of the textile insulation with its moisture content which in turn varied with the surrounding humidity. It was found that the application of a thin film of cellulose acetate to the textile insulation considerably stabilized its properties.

The action of the cellulose acetate coating is illustrated in Fig. 1. When the surface of untreated wire is magnified, it is evident that it is a mass of extending fibers and when two conductors lie adjacent, these fibers interlace. Under humid conditions the textile becomes moist and the interlacing fibers afford a path for current leakage. The moist fibers also have a considerably higher dielectric constant than the dry fibers and the electrical capacitance between adjacent wires is increased, therefore, with high humidity. In the lower half of the figure, the treated wire is shown, with the fibers ironed and sealed down by the cellulose acetate process. It is not claimed that the film is free from cracks, for cracks do form in the subsequent twisting and forming operations permitting moisture to enter, but no matter how moist the textile may become, no direct path for the leakage of current is formed as the fibers are separated from one another by two films of cellulose acetate.

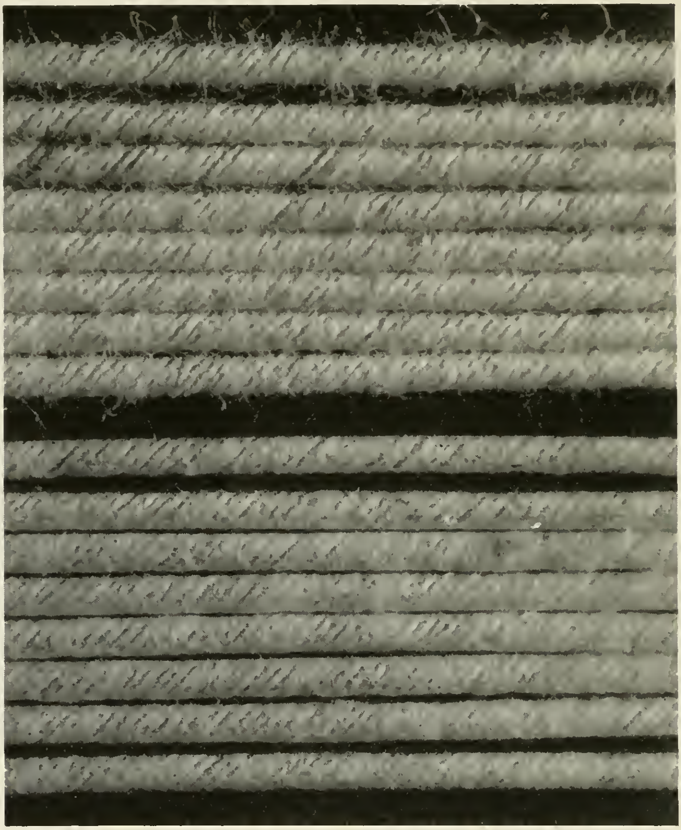


Fig. 1—Textile insulated wire before and after treatment with cellulose acetate. (Enlarged 4 diameters.)

Development of Process

Extrusion of a coating as thin as wanted, about 0.0015 inch, appeared to be out of the question. Even if it were possible the coating would not adhere firmly enough to prevent cracking off when the wire is twisted or otherwise roughly handled. The study of method was therefore confined to coating the textile insulated wire from a solution. Several solvents for the flake cellulose acetate were considered but acetone was adopted, at least for the time being, as it was known to have the desired characteristics and would simplify the problems involved. The solution would have to be low in viscosity to produce the thin coating desired and therefore the quantity of acetone used would be high. This made it economical to recover the acetone after it had been driven off from the coated wire.

It was quickly established that the procedure should include several passes of the wire through the cellulose acetate solution with the excess solution removed and the coating dried after each pass. The source of drying heat was optional and for experimental purposes electrical heating elements were used. The temperature of drying was limited by the tendency of the coating to blister. A rapid circulation of air quickened the rate of drying and with a sufficiently long drying chamber, a high rate of wire travel seemed possible.

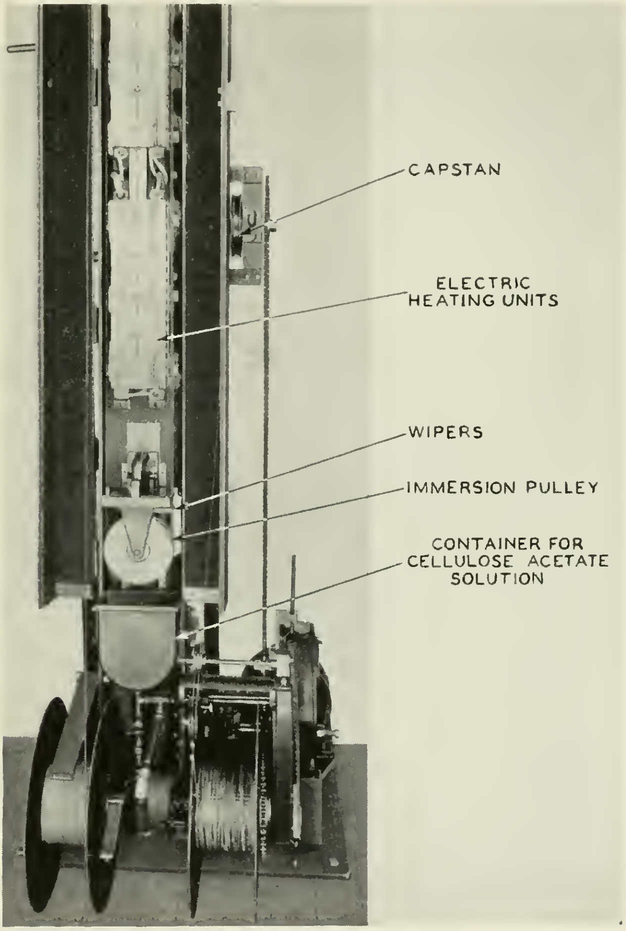


Fig. 2—An experimental machine for treating textile insulated wire with cellulose acetate. The immersion pulley with wiper dies are shown raised for threading.

An early experimental machine is shown in Fig. 2. At the bottom are the supply reel and the motor driven take-up reel. The cellulose acetate solution is in the small tank near the bottom The wire passes around the immersion pulleys, shown raised for threading, then past wipers into the drying chamber, over an upper pulley, and down, for a total of four passes, which was later increased to six passes.

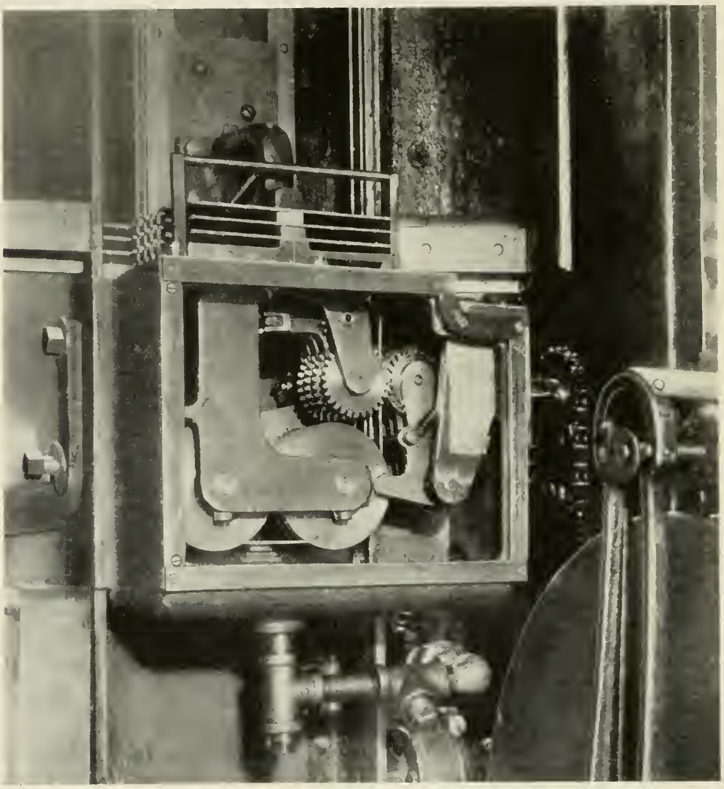


Fig. 3—A later experimental machine, equipped with rotating dies for removing the excess cellulose acetate and pressing down the textile fibers.

It was found that stationary wipers quickly became clogged by a mixture of lint from the textile serving and half dried cellulose acetate solution, and the rotating dies shown in Fig. 3 were designed. These dies are the heart of the machine and their construction and adjustment were found to be critical. It was necessary that they should allow wire splices and enlarged places in the textile insulated wire to pass, and to permit this the right hand dies were carried on spring arms allowing them to back away. They were driven in the same direction as the wire travel but at one-third the speed. The slower speed causes a wiping action on the wire, ironing down the fibers, but still tends to feed wire splices through the dies. The size of the grooves must be proportioned exactly for each gauge of wire, the first pass being the largest. This is necessary so the extending fibers will be gathered in rather than caught between the dies and cut off. During the next three passes these fibers are gradually ironed down, as illustrated in Fig. 4, while the fifth and sixth passes are largely for smoothing down and sizing. Adjustments were provided for lining the dies up with the wire so the pressure on both sides of the wire would be equal, resulting in an even coating. Means were provided

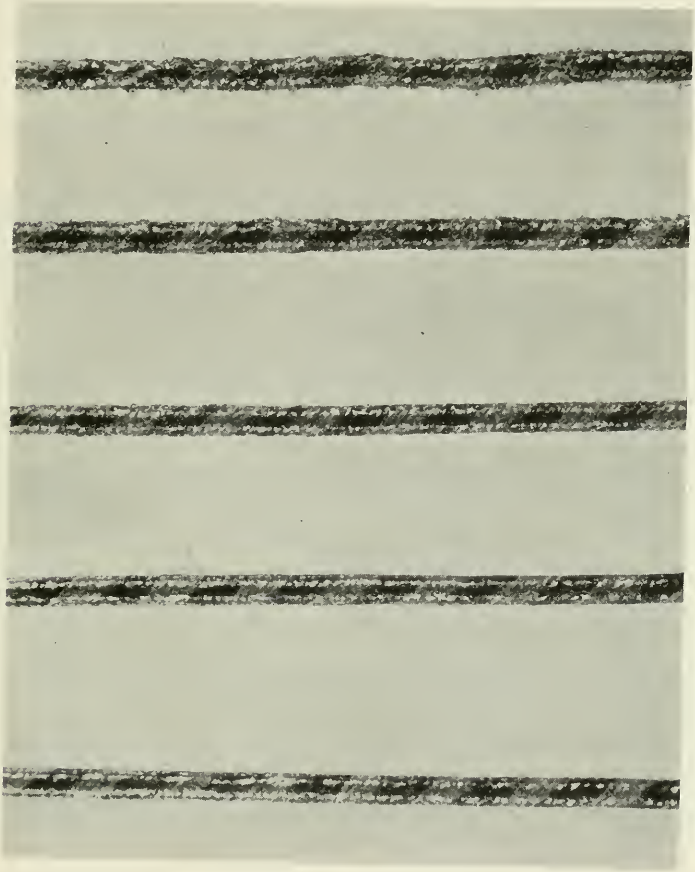


Fig. 4—An enlargement showing the gradual pressing down of the textile fibers and smoothing of the cellulose acetate coating as the wire receives the successive applications.

for quickly removing the dies for cleaning as it was expected that soaking them in acetone would be necessary at frequent intervals.

The density and viscosity of the cellulose acetate solution have an important bearing on the process. It was desirable to use as little acetone as possible as this would hasten the drying operation and with less acetone in the process, less acetone would be lost.

To obtain the maximum effect from the film of cellulose acetate, it was desirable that the coating be as free from extending fibers and as smooth as possible. Passing the coated wire while still plastic over pulleys so arranged that the entire circumference of the wire would come in contact with the pulleys, was found to produce a polishing effect.

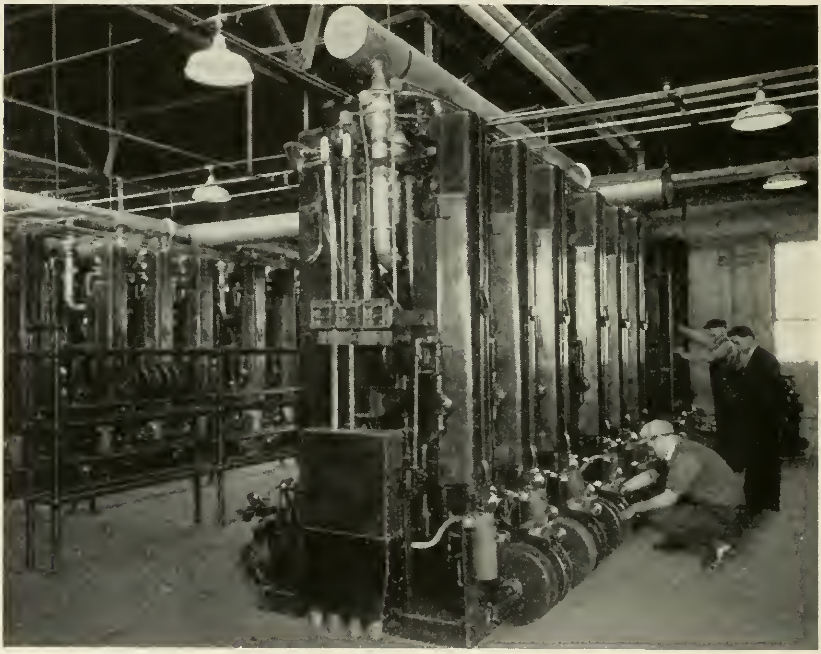


Fig. 5—An installation of machines for treating textile insulated wire with cellulose acetate.

Commercial Installation

With the essentials of the cellulose acetate coating process determined, the design of a commercial machine was undertaken. Figs. 5 and 6 show a number of these machines which, it will be seen, are in many ways similar to the experimental machine previously illustrated. Each head is separately driven by a two-speed electric motor and there is a second motor for circulating the air within the head. The main drive motor is connected so that it rotates the wiping dies continuously so they will not freeze up, and the take-up reel and the capstan are operated through a hand clutch. Six passes of wire were decided upon as giving better control of the coating than afforded by four passes. Electric heating elements were selected as being compact and easily controlled. In the schematic view, Fig. 7, are shown the method of air supply, the circulation of the air within the drying chamber, and the exhausting of the acetone-air mixture.

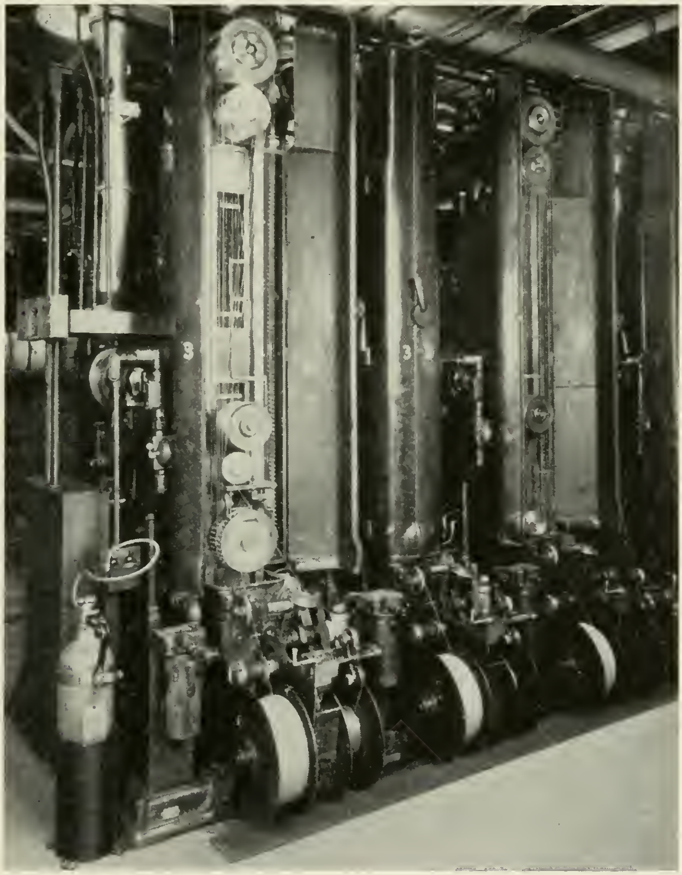


Fig. 6—Cellulose acetate wire coating machine.

In making a commercial installation of the new machines, it was necessary to go into the questions of preparing and circulating the cellulose acetate solution, recovering the acetone, and making the entire process as safe as possible from the hazards that are always present when using a solvent such as acetone.

Fig. 8 illustrates the general arrangement of the equipment. Two buildings are used instead of one to separate the large quantities of solution present in the mixing room from the room in which the coating machines with their electrical equipment are located. The acetone is pumped from the underground storage tanks to the mixers. The cellulose acetate is weighed out from bins and added to the mixers and the whole stirred for several hours until fully dissolved. One gallon of acetone is used to ten ounces of cellulose acetate. The solution is then pumped continuously from the mixers to the machines. From the machines it flows to a sump by gravity and from the sump is forced back to the mixers. The reasons for circulating the cellulose acetate solution are to keep it homogeneous, to keep the amount of solution in or near the machines as small as possible, and to permit control and maintenance of its viscosity at one location instead of in the individual dope pots.

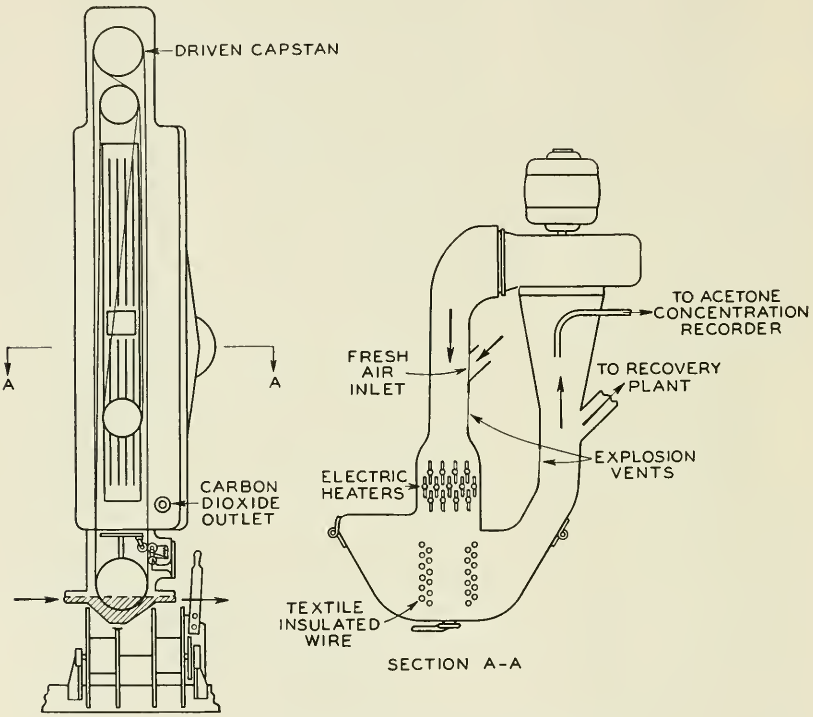


Fig. 7—Schematic view of one coating head.

It was found that the viscosity of the solution had to be maintained within close limits, as the weight of film left on the wire changes with the viscosity. To continuously indicate the viscosity, a constant amount of solution is pumped through an orifice and the back pressure developed is measured by a pressure gauge, the gauge being calibrated for viscosity. Deviations from the required viscosity are corrected by the addition of acetone or a more concentrated solution that is kept in readiness.

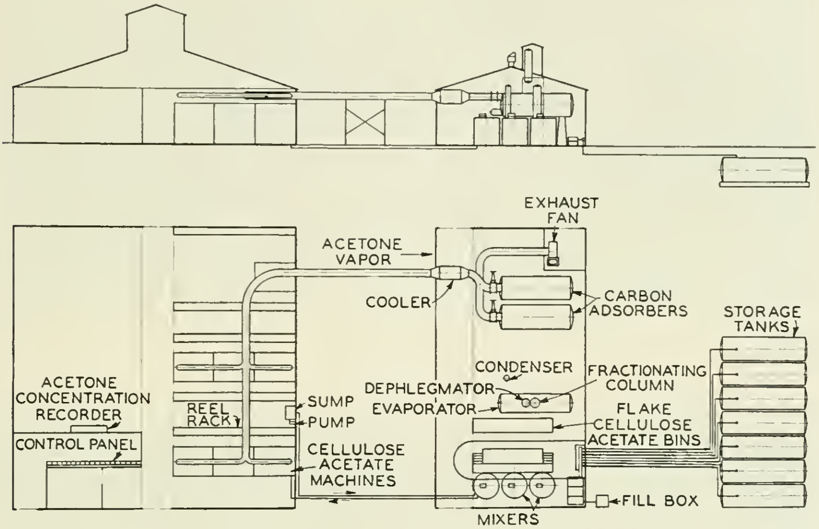


Fig. 8—Schematic of the wire treating plant at the Kearny, New Jersey works of the Western Electric Company.

To keep the amount of acetone solvent low for the required viscosity, the temperature of the solution is kept at about 86° F. Although a higher temperature would require the use of even less solvent, it would increase the loss by evaporation. To maintain the desired temperature, a water coil was installed in the mixer, the water in turn being heated by steam. This indirect method was used instead of installing steam coils directly in the mixer as the latter would result in boiling the cellulose acetate solution and the coils would become heavily coated with cellulose acetate. The mixers are equipped with standard propeller type agitators driven through gear reduction by totally enclosed motors.

During the drying operation the acetone is evaporated giving a mixture of air and acetone vapors within the drying chambers. This mixture would be explosive if the concentration of acetone vapors were

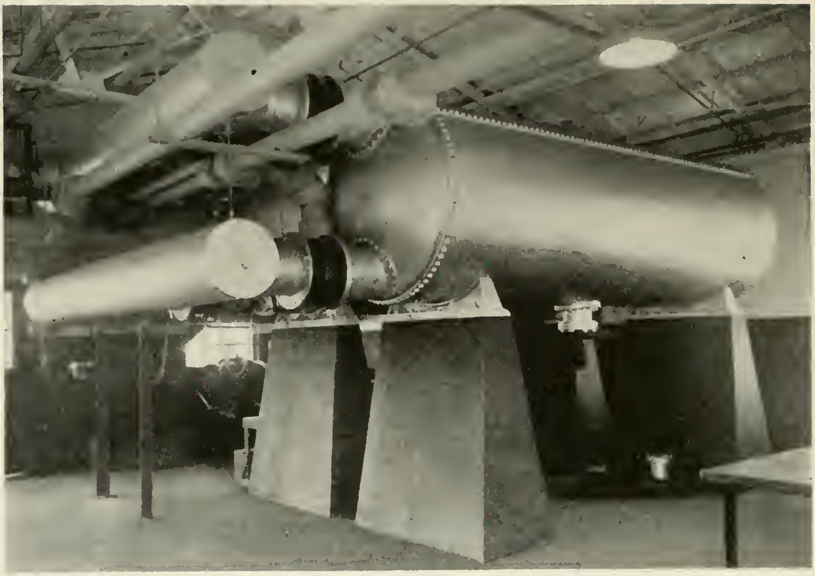


Fig. 9.—Charcoal adsorbers for removing the acetone vapors from the exhaust from the treating machines.

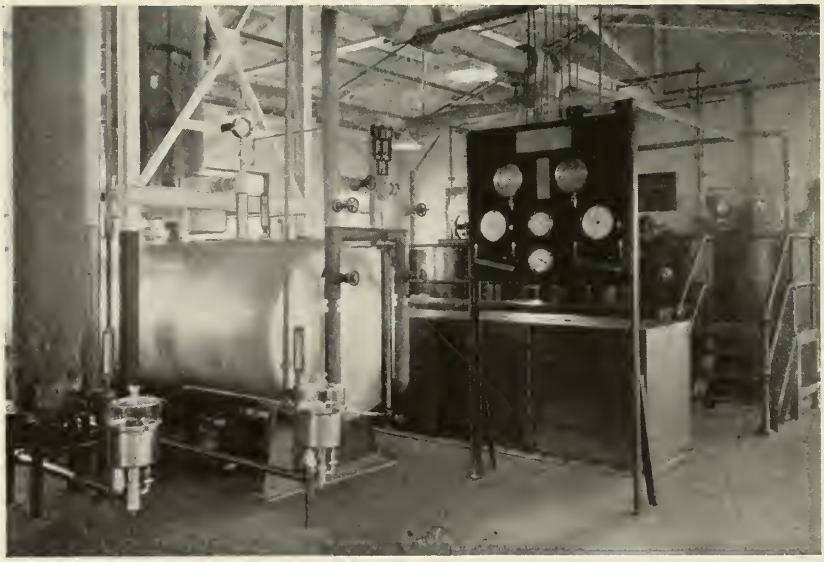


Fig. 10—Equipment for the distillation of acetone.

allowed to get between approximately 2.5 per cent and 13 per cent by volume and at the concentration of 5 per cent an explosive force of as high as 75 pounds per square inch might result. To eliminate the possibility of a vapor explosion, enough of the mixture is constantly drawn away through an exhaust duct so that the concentration of acetone will not exceed 1.5 per cent. The vapors are conducted back to the building in which the mixing was performed and the acetone is recovered.

The acetone recovery system shown in Figs. 9 and 10 consists of activated carbon adsorbers, a condensing tower, and a dehydrating plant. The air and acetone mixture coming through the exhaust header from the coating machines passes through a water cooler to the carbon adsorbers where the acetone is adsorbed and the air is exhausted through the roof. After about an hour's operation, the carbon in one of the tanks has adsorbed as much acetone as it can without allowing a portion of it to pass through and the air valves are switched to pass the acetone air mixture through the other adsorber. Steam at low pressure is then admitted to the first adsorber and the acetone is driven off. The steam and acetone vapors are condensed, giving a condensate that averages about half water and half acetone. This flows by gravity to one of the outside underground storage tanks where it is held until it can be dehydrated. When a supply of water-acetone has accumulated, it is pumped to the evaporator and the dehydrating is performed in the usual way. The recovered acetone flows by gravity to another of the underground tanks for reuse. Samples are taken at intervals to insure that the dehydration is complete, that no foreign materials are present and that no breakdown of the acetone has occurred. More than 85 per cent of the acetone purchased is recovered.

There were, of course, other types of recovery systems available, but for our purpose the carbon adsorption process was most economical. Scrubbing with water or other absorbing mediums would not give as high a percentage of recovery and would be more expensive to install and operate. Freezing out the acetone would not be as economical with the low concentration of acetone-air mixture that it was felt must be maintained to afford safety from explosion.

Precautions Against Fire

Throughout the engineering of this installation, the greatest precautions were taken against possible fire and explosion. Double forty-mesh screens were installed close to the junction of the vent pipes with the underground tanks. The mixing tanks were equipped with explosion reliefs consisting of .002 inch thick sheet aluminum. The tanks were vented to the recovery system, with a fire screen installed in the vent line. Attached to the mixing tanks is a carbon dioxide fire extinguishing system. One opening is within the tank and another one is above the explosion relief. This carbon dioxide system will be released automatically if a quick rise in temperature occurs within the mixer and it can also be operated through a hand pull box on the outside of the building. Operation of the carbon dioxide system in the mixers or in any other place in the cellulose acetate plant causes all electric power in both buildings to be turned off instantly, sounds a siren and calls the fire department.

To avoid the possibility of static sparks while loading the flake cellulose acetate into the mixer, a humid atmosphere is obtained by injecting steam. The danger is further lessened by the avoidance of metal in the container used for carrying the flake acetate.

In the coating machines themselves, precautions have been taken both to prevent fires and to render them harmless if they should occur. A temperature well above 1,000° F. is required to ignite acetone and the electric heating elements are operated at about half this temperature. The temperature of the drying chamber is controlled by a pyrometer which turns on or shuts off the electric heating current as required. As a further safeguard, a second temperature control, set at a slightly higher temperature, is arranged to act if the first one fails, shutting off the current and sounding a gong. Here again a carbon dioxide fire extinguishing system is permanently attached with an outlet in each drying chamber. Automatic discharge would be difficult to control because of the temperature conditions in the chamber and the carbon dioxide is controlled entirely from hand pull boxes. Hand CO_2 equipment is also available in convenient locations. Explosion vents covered with aluminum foil are located in the rear of each unit.

However, as previously mentioned, there is no danger of explosion of the acetone-air mixture in the drying chamber if the percentage of acetone is maintained at less than the amount that will propagate flame, about 2.5 per cent. Acetone-air analysis instruments of a standard make are permanently installed, taking readings from each of the drying chambers and from various points in the exhaust lines. These instruments operate on the basis of difference between the thermal conductivities of air and of atmospheres containing various amounts of acetone vapor. Two platinum coils, heated by a constant electric current, form two arms of a Wheatstone Bridge. Around one of these coils is room atmosphere, containing whatever moisture and other impurities it may, while around the other coil are the gases drawn from the drying chamber, cooled down to the same temperature, containing acetone vapors in addition to the constituents of the room atmosphere. The different thermal conductivities of the gases around the two coils cause a difference in temperature and therefore a change in resistance of the platinum coils which causes a reading on the galvanometer. The instrument is calibrated to read directly in per cent acetone. If a reading greater than the allowed maximum of 1.8 per cent concentration is encountered, a gong is sounded and the head is immediately switched to half speed until the trouble is located and remedied. The percentage of acetone concentration can be varied by increasing or decreasing the amount of room air drawn through the machine to the recovery system.

To avoid all sparks in the main rooms, where some acetone vapor may exist, a separate control room with entry from the outside only is provided. Lead covered cables running through sealed ducts connect the machines with the control room and all fuses and relays are in this room. Switches on the machines are oil immersed or fully enclosed and gas tight. A magnesite floor is provided to prevent the striking of sparks from reels or tools that may be dropped and brass floor plates are used instead of steel for the same reason.

Although the present installation is well protected from fire and explosion, there are still a number of characteristics of acetone that are not well known, and it is probable that more exact knowledge would make it possible to achieve safety in future installations at less expense. A study is therefore under way in cooperation with the Bureau of Mines at Pittsburgh to determine these unknown characteristics of acetone. Among the unknown quantities are the exact lower limit of acetone vapor concentration that is explosive under conditions such as are encountered in our drying chambers, the effectiveness of various relief openings, the probability of ignition from various sources, especially static sparks, and the force of explosions possible from various concentrations.

Recent Developments

A new coating machine has been developed that operates at a higher speed, is simpler to operate and is still better safeguarded against fire and explosion. It consists of eight coating units on a common base, four units being on each side. Instead of electrical heating elements in each coating unit one steam heater supplies hot air to a number of machines. The temperature is automatically controlled at the heater, leaving the operator free to watch the supply and take up of the wire. A wire speed of 240 feet per minute has been obtained compared to

150 feet per minute with the electrically heated machines. With an overall efficiency of 90 per cent, the production per head is 13,000 feet per hour.

Ample explosion reliefs have been included in the new machines. For the chamber itself, a diaphragm of aluminum foil that will rupture leads to an open space between the machines. On the top of the air ducts above the machines, vents are located that will relieve any explosive pressure in this part of the system before it reaches dangerous proportions.

Although the coating of textile served wire with cellulose acetate is being performed successfully, it is felt that the ultimate design of machine or maximum efficiency have not been reached. There is promise in the possibility of using more viscous solutions of cellulose acetate, thus reducing the amount of acetone to be evaporated and recovered. Other solvents than acetone are to be tried. Higher machine speeds may be found possible and operating methods will be developed that will increase the uniformity of the product.

The Development of a Handset for Telephone Stations*

By W. C. JONES and A. H. INGLIS

A number of factors contribute to the difficulties involved in the design of a telephone handset which gives as good service performance as a deskstand. The handset transmitter, for example, not only is used in a wider range of positions but also is moved much more frequently, so that wider variations are experienced in its characteristics.

Further difficulties are introduced by the close physical connection of the receiver and transmitter, in that "howling" tends to be set up.

The handset has been developed so that it overcomes all these difficulties and is interchangeable with the deskstand in existing telephone plant without important reaction on either transmission or signaling performance.

MANY interesting development problems are presented in the design for general use of a handset which provides the convenience of this arrangement of a telephone set without a sacrifice in the performance of the system. It is the object of this paper to discuss some of these problems and to describe their solution as embodied in the handset now being furnished by the Bell System.

The idea of mounting a telephone transmitter and receiver on a common handle to form a handset was conceived early in the development of the telephone. In 1878, only a few years after the invention of the telephone, handsets of the type shown in Fig. 1 were in use by operators in the Gold and Stock exchange in New York City.[1] Variable resistance transmitters of the Edison type were used in these handsets. This transmitter employed a relatively insensitive lamp black resistance element which was soon superseded by the more sensitive granular carbon type in order to permit the extension of telephone service to greater distances. The basic ideas underlying the variable resistance transmitter and the many advantages of granular carbon over the numerous other materials which have been tried have been discussed elsewhere.[2] It is sufficient to point out here that the large amplification afforded at low cost by a well designed granular carbon transmitter makes it unlikely that any other structure will offer successful competition in general telephone application for some time to come.

When an attempt was made to use granular carbon transmitters with handsets it soon became evident that a satisfactory design involved more than the mounting of the available types of transmitters and receivers on a common handle and that considerable development of the instruments as well as the means for coupling them would be required before a handset suitable for general use were obtained. For this reason, the telephone system in this country has in general been built up around the wall set and the deskstand which permitted the utilization of the available transmitters and receivers to best advantage.

---

* Presented at A. I. E. E. Midwinter Convention, Jan. 25–29, 1932, New York, N. Y.

[1] "Beginnings of Telephony," F. L. Rhodes, Harper and Brother, 1st Edition, p. 153.

[2] "The Development of the Microphone," H. A. Frederick, *Journal of the Acoustical Society of America*, July 1931, Part 2, p. 17; *Bell Telephone Quarterly*, July 1931.

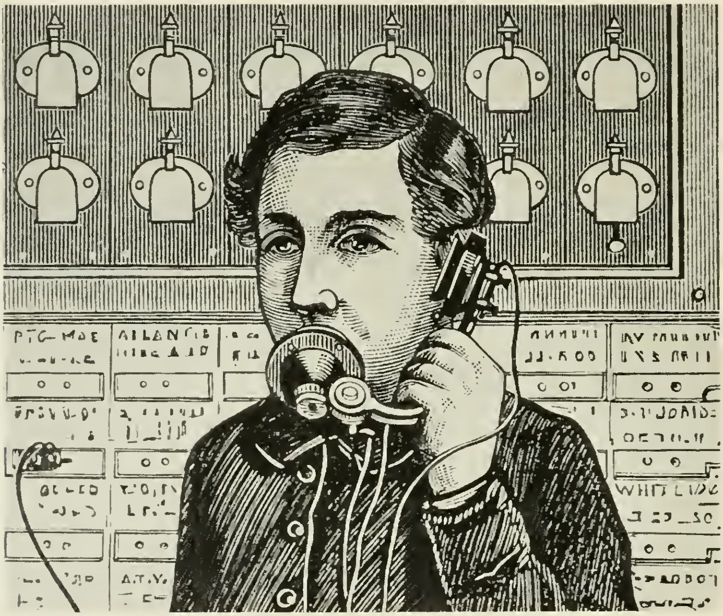


Fig. 1—Operators handset used in 1878.

A number of factors contribute to the difficulties involved in the design of a handset which gives as good service performance as a deskstand. Some of these are due to a difference in the conditions of use of the two instruments and some are due to the necessary differences in structure. A consideration of these differences will afford a background for the subsequent description of the line of attack followed in providing a satisfactory handset.

Difference in Usage

The controlling differences in service conditions are due to the fact that the handset permits the user much more freedom than the deskstand. The handset transmitter, accordingly, is not only used in a wider range of positions but is also moved much more frequently with the result that, in general, much wider variations are experienced in its transmission characteristics and resistance. If the transmitter is not suitably designed these variations may be sufficiently large to render the transmission of a handset unsatisfactory, and to interfere with the operation of associated relays and other signalling apparatus.

The more severe use to which the handset is subjected also has a tendency to materially accelerate changes in the contact surfaces of the carbon and the electrodes. These changes are evidenced principally by increased carbon noise and resistance which affect both transmission and signalling and appreciably shorten the useful life of the transmitter unless proper design measures are taken to reduce these aging effects.

Differences in Structure

It is undesirable in any telephone set, designed for use by the general public, to require the user to perform any switching operation in changing from talking to listening. For this reason, practically all commercial telephone sets are of the so-called "invariable" type, in which the transmitter and receiver are at all times connected to the line while the set is being used.[3] Part of the output of the transmitter, therefore, is transmitted to the receiver of the local telephone set. This electrical connection between the transmitter and receiver of the same set is known as the "sidetone" path, and the resulting sound in the receiver as "sidetone." In addition to this electrical coupling between the instruments, there is acoustical coupling through the air and in the case of the handset mechanical coupling through the handle. If the amplification afforded by the transmitter is greater than the aggregate losses in the sidetone path, the receiver, the air and the handle, sustained oscillation or "howling" may be set up. Any such condition, is, of course, not only unpleasant but fatal to transmission. It has been found necessary, in fact, to keep well below the howling point to avoid serious transmission impairment due to transient oscillations.

In the case of the deskstand, the sidetone path and the air path provide the only coupling between the instruments and the losses are of sufficient magnitude that howling does not occur even with the most efficient transmitters and receivers. In the case of the handset, however, the handle may add appreciably to the coupling. Without proper design of the handle, instruments, and means for mounting the instruments on the handle this additional coupling is often sufficient to cause howling.

---

[3] "Transmission Circuits for Telephonic Communication," K. S. Johnson, D. Van Nostrand Co., 4th printing, p. 105.

Since with the deskstand a flexible cord forms the only connection between the transmitter and receiver, the user is always able to hold the receiver to his ear and at the same time to speak with his lips directly in front of the transmitter. The handset handle, however, definitely establishes the distance between the two instruments. If it is too short, many users are unable to hold the receiver on the ear when the transmitter is in front of the lips. This difficulty has often been avoided by making the handle long enough to accommodate any user. If this is carried to an extreme, the distance between the transmitter and the lips is greater than necessary, and introduces an avoidable transmission loss for the majority of users. As may be seen from the curve shown in Fig. 2, this loss may be quite large and warrants every effort to minimize it.

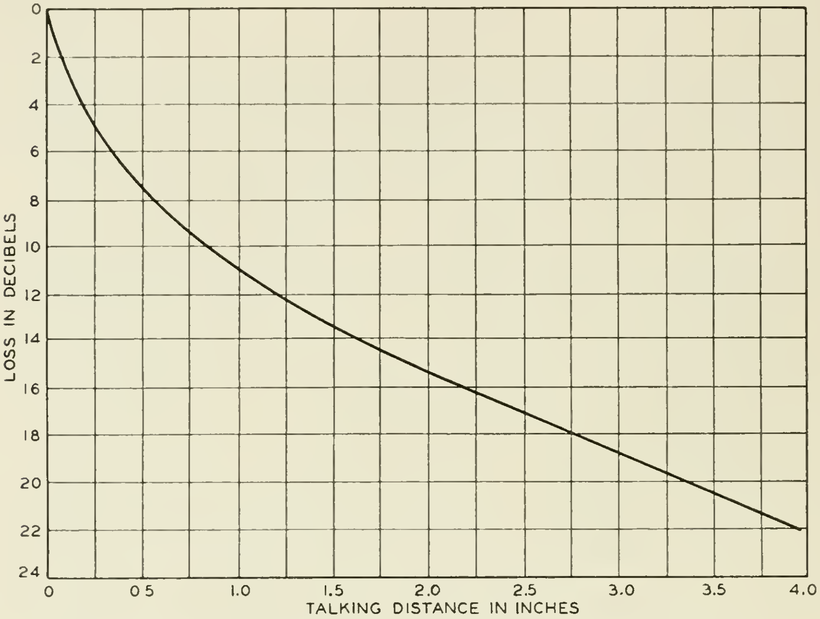


Fig. 2—Loss in transmitter output with distance from the mouthpiece.

If the convenience of the handset is to be fully realized, it is important that the handle be shaped to fit comfortably in the hand and that the complete handset be sufficiently light to avoid fatigue on the part of the user. These structural limitations add a further problem in design.

DESCRIPTION OF THE HANDSET

While many incidental problems have required solution during the development of the present handset, which is shown with its mounting in Fig. 3, this design is largely the result of a systematic attack on the more fundamental problems that have been discussed. In describing the handset, therefore, particular attention will be given to those features which are of importance in the solution of these problems. Some of the most important and interesting of these are embodied in the transmitter.

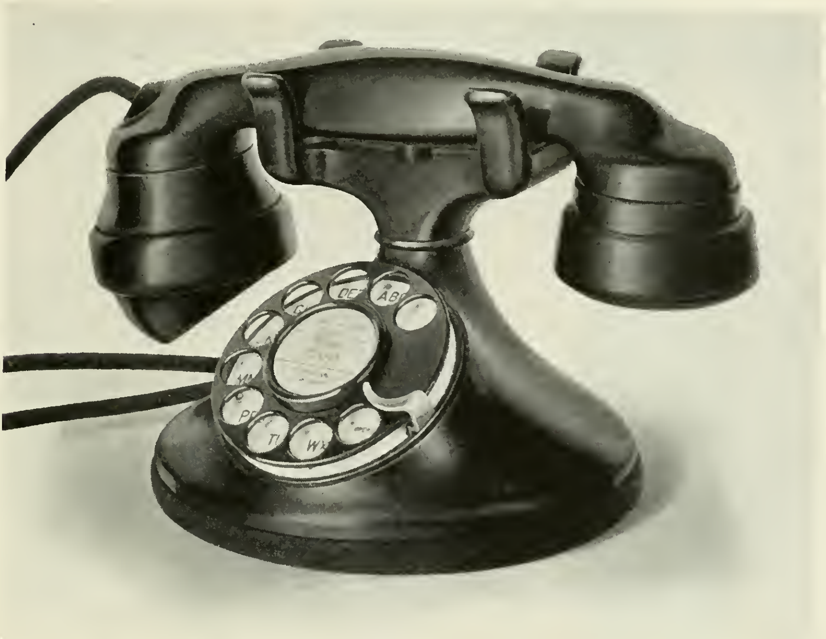


Fig. 3—Station handset and its mounting.

TRANSMITTER

Referring to the cross-sectional view of the handset, Fig. 4, it will be observed that the various parts of the transmitter are assembled in a die cast aluminum housing to form a unit which mounts in a threaded brass bushing in the handle. One electrical connection is made through this bushing, the other through a contact which engages a spring in the base of the housing. The molded phenol plastic mouthpiece, dome and spacing ring, which form the external parts of the

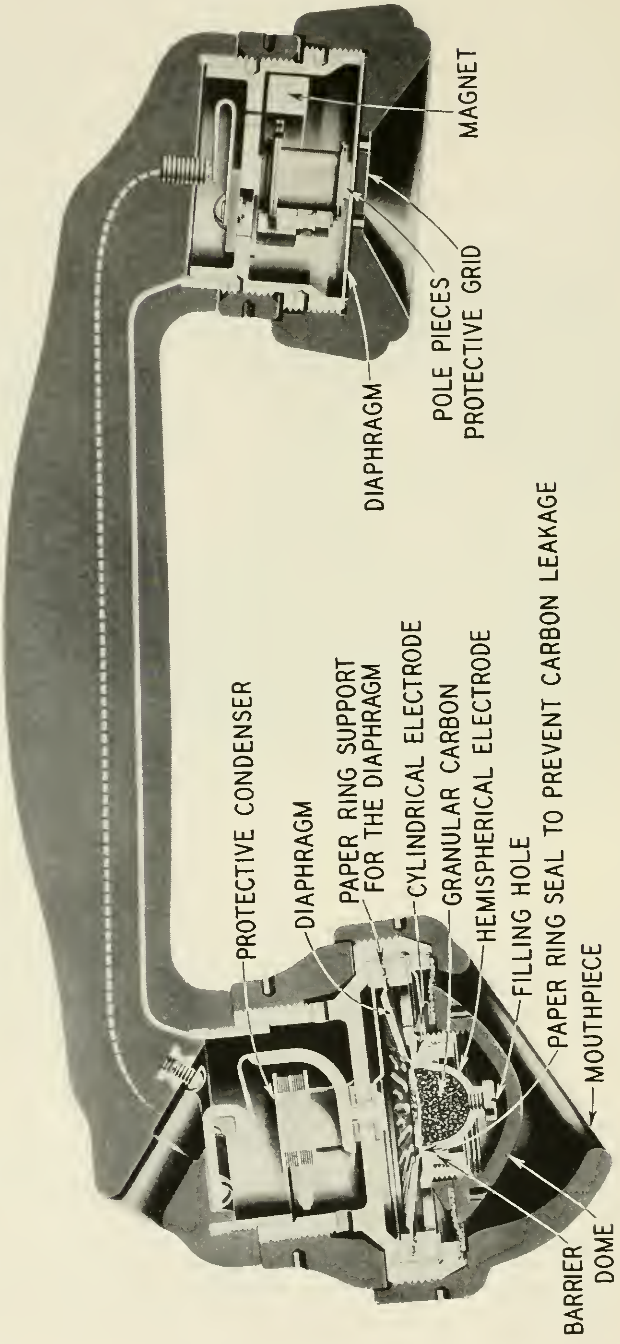


Fig. 4—Cross section of the station handset.

transmitter, insulate the metal parts in the electrical circuit so that they cannot come into contact with the user. The spacing ring also serves to lock the transmitter to the handle and align the mouthpiece with the axis of the handle.

Effect of Angular Position on Resistance

The carbon chamber of the transmitter is often referred to as a "barrier" type and differs radically from the conventional form of "direct action" resistance element, which has been employed for years in deskstand transmitters. In the direct action type the movable member not only transmits the acoustic forces on the diaphragm to the carbon but also serves as an electrode. Both electrodes of the barrier type are stationary and a thin layer of phenol varnish insulates the diaphragm from the carbon. A ceramic barrier [4] separates the electrodes from one another and defines the current path through the carbon. The electrode surface adjacent to the diaphragm is cylindrical in shape, the other hemispherical. Both are gold plated. By adopting this electrode arrangement good contact is maintained be-

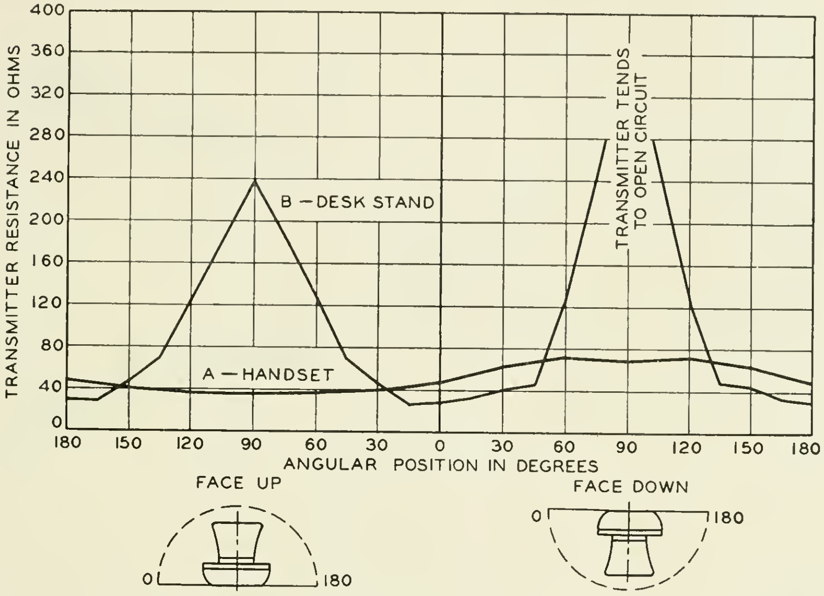


Fig. 5—Effect of position on transmitter resistance.

---

[4] "Manufacture of Thin Porcelain Parts of Close Dimensional Tolerances," L. I. Shaw, A. O. Johnson and W. J. Scott, *Journal of Ceramic Society*, Nov. 1931, pp. 851–854.

tween the carbon granules and the electrode surfaces and between the granules themselves, in all of the positions in which the transmitter is likely to be held while in use. This uniformity of contact pressure makes the resistance of the transmitter nearly independent of angular position as shown by Curve *A*, Fig. 5.

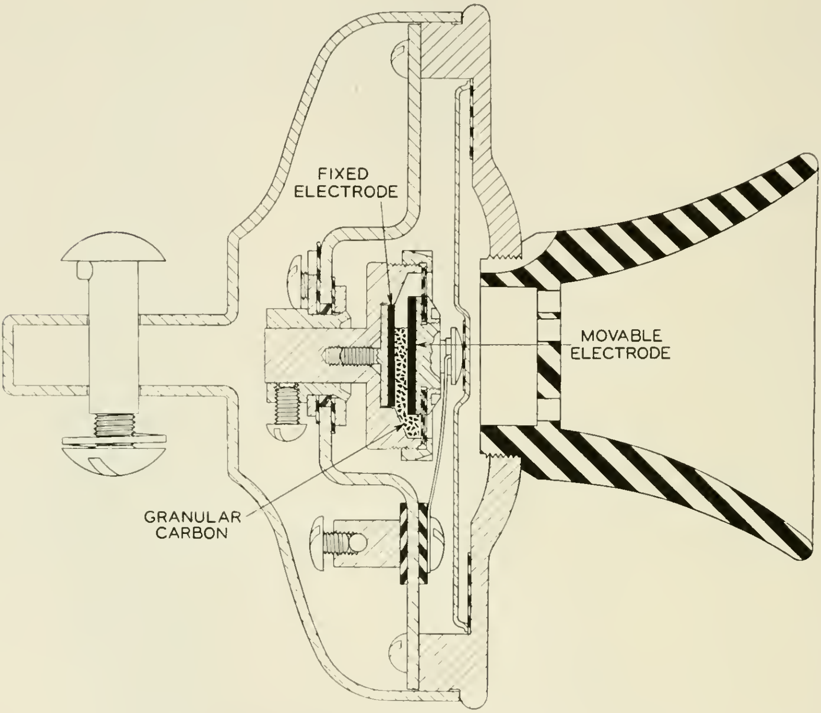


Fig. 6—Cross section of a typical deskstand transmitter.

The substantially uniform resistance of the present handset transmitter in various positions represents an outstanding improvement over that obtained with a typical deskstand transmitter such as that shown in cross-section in Fig. 6. Referring to curve *B*, Fig. 5, it will be observed that the resistance of this transmitter is markedly dependent upon angular position, two prominent peaks occurring where the electrodes are practically horizontal and the carbon falls away from the surface of the upper electrode. This is typical of the performance of the older types of transmitters employing the conventional form of direct action structure. While this characteristic is not important in deskstand service it renders this type of transmitter unsuitable for the more severe conditions encountered in the use of a handset.

Effect of Angular Position on Carbon Noise

Another transmitter characteristic which is likely to be influenced by the position in which the handset is held is carbon noise. In addition to the adsorbed gas on the surface of the granule there is also an appreciable amount of gas absorbed in the pores in its surface.[5] When a voltage is impressed on the transmitter the temperature [6] of the points of contact between the granules is increased and a certain amount of the gas is driven off, forcing the granules apart and giving rise to non-periodic changes in resistance. These disturbances result in noise in the receiver which tend to mask incoming speech. It is evident that a given efflux of gas will produce a larger change in resistance when the contact forces are low than when they are high. Hence, if as a result of turning the transmitter through the various angles in which it is likely to be used, the contact pressure is lowered and the resistance raised, a sufficient increase in temperature may occur to cause a vigorous evolution of gas and increased noise. It would, therefore, seem reasonable to expect that, inasmuch as the contact

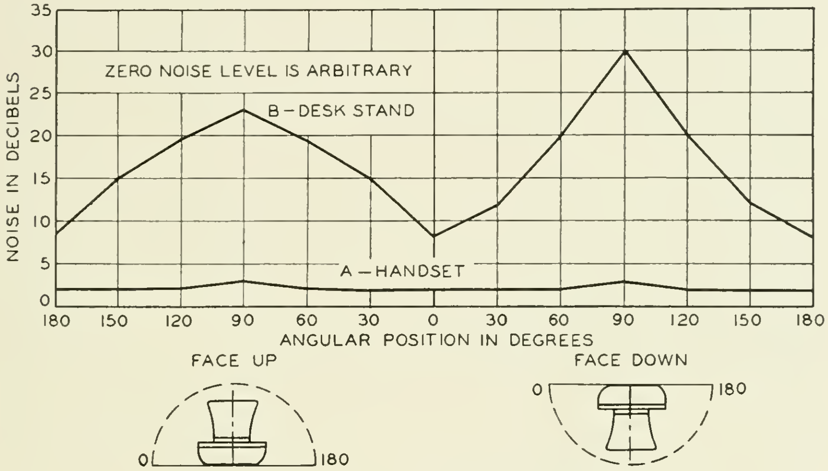


Fig. 7—Effect of position on carbon noise.

---

[5] "The Effect of Gases on the Resistance of Granular Carbon Contacts," P. S. Olmstead, *Journal of Physical Chemistry*, Jan. 1929, pp. 69–80.

[6] "Über Kontaktwiderstände besonders bei Kohlekontakten," Von Ragnar Holm, *Zeitschrift für Technische Physik*, Sept. 1922, pp. 290–294, Oct. 1922, pp. 320–326, and Nov. 1922, pp. 349–357.

pressures and the resistance of the handset transmitter are substantially independent of angular position, the change in noise with position would be practically negligible. That this condition is realized is evident from the data shown in Fig. 7. Not only is the average carbon noise negligible even under severe conditions of service, but the maximum variation of the noise in any position of the transmitter is only about one db. On the other hand, the noise developed by a typical deskstand transmitter in the optimum position is approximately five db higher, and in its worst position nearly thirty db higher than the noise produced by the handset transmitter. Noise of such a magnitude would constitute an appreciable transmission impairment.

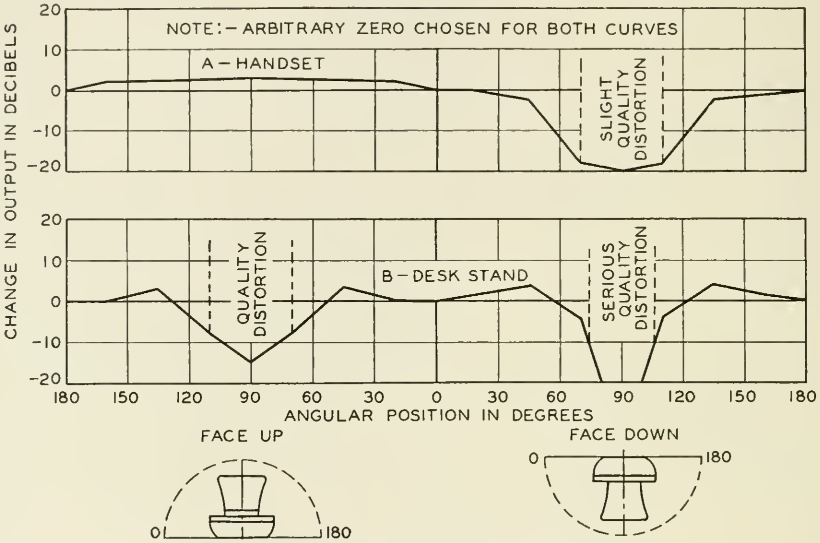


Fig. 8—Effect of position on transmitter output.

Effect of Angular Position on Output

One of the unique features of this new handset transmitter is the fact that the carbon chamber is located in front rather than in back of the diaphragm as has been customary in the past. By adopting this arrangement the carbon granules are held in intimate contact with the diaphragm in all of the positions in which the handset is likely to be held and uniform output and faithful reproduction of the speech sounds obtained. There are, however, angular positions in the region of 90 degrees face down where the carbon tends to fall away from the diaphragm, but as is shown by curve *A*, Fig. 8, the resultant loss in the output and the degradation of the quality are limited to a rather narrow range of positions in which the handset is seldom held. A loss in output and a serious impairment of quality occurs when the usual type of deskstand transmitter is held in the positions which would frequently occur if it were used for handset service. This effect is shown by curve *B*, Fig. 8.

Response

The diaphragm of the transmitter is made from thin duralumin formed into a truncated cone with radial stiffening ribs. This reduces the effective mass to about one-tenth that of the deskstand transmitter and provides sufficient rigidity to insure vibration as a unit throughout the frequency range of interest. A number of impregnated paper rings each approximately four ten-thousandths of an inch in thickness support the edge of the diaphragm. The dimensions of the recess into which these rings assemble are so chosen that they separate slightly from one another. This construction provides a resilient support for the diaphragm and adds a certain amount of mechanical resistance due to the viscosity of the air films between adjacent layers of paper and the friction between the layers. An appreciable improvement in response results from the lower mass and stiffness and the higher damping. The range in response of the handset transmitter in the frequency range from 300 to 3000 cycles per second, is about 20 db as compared with approximately 40 db for the deskstand transmitter over the same frequency range. The more uniform response of the handset transmitter causes a marked improvement in articulation over that obtained with the deskstand instrument.

Aging

As previously mentioned, after a granular carbon transmitter has been in use for a time, changes take place which often cause an increase in resistance and carbon noise and a loss in sensitivity. These changes in the contact conditions have been traced to two principal causes, the abrasive action of one granule on another, and the deterioration of the surface of the granules due to high temperatures resulting from excessive contact voltage. The former is usually referred to as "mechanical aging," the latter as "electrical aging."

The aging of the deskstand transmitter is primarily electrical, for although it is picked up occasionally and moved from one location to another on the desk or table, it is not in general subjected to shocks or jars which cause an appreciable motion of the carbon granules. Because of the more severe service conditions, the predominant factor in the aging of the handset transmitter is mechanical.

Laboratory aging tests equivalent to about four years of handset service under severe conditions, indicate an increase in carbon noise of only 6 to 8 db above the initial value. Neither before nor after aging is the noise produced by the handset transmitter of practical importance.

The output of the handset transmitter is decreased somewhat and the average resistance increased to about double its initial value as the result of aging. The relatively small change of resistance with position in this design, however, effectively prevents the frequent occurrence of resistances sufficiently high to interfere with the operation of signalling apparatus in circuit with the transmitter. In this respect the new design represents a notable improvement over earlier types of instruments.

The method adopted for filling the handset transmitter contributes materially to keeping the aging rate low. In most of the transmitters which preceded the present instrument, the carbon occupied only about three-quarters of the volume of the carbon chamber. This allowed the granules to move freely and caused rapid aging. Obviously, the movement of the granules would be reduced to a minimum if the chamber were filled full of carbon. Inasmuch as the space occupied by a given weight of granular material is dependent upon the configuration of the granules, it is evident that a full filling cannot be obtained merely by pouring carbon into the chamber. Definite means, therefore, must be provided for bringing the carbon into a minimum volume state if a full filling is to be obtained.

A machine has been designed for this purpose which vibrates the transmitter during the filling operation. This settles the carbon in much the same way as if the transmitter were held in the hand and tapped. The carbon chamber cannot only be filled full of carbon in this way, but the effect of differences in the volume of the chamber, due to the commercial variations in the size of the parts, is eliminated.

It has been found that the loss in sensitivity which results from filling a transmitter in this manner is relatively small. There is, however, one feature of a full filling, often referred to as "mechanical packing," which is objectionable and has led to a filling slightly less than full. If the diaphragm of a transmitter which has been filled full by the method described, is displaced sufficiently to permit the granules to change their configuration, the volume of the filling will increase and the diaphragm will not return to its original position. A marked increase in contact pressure and lowered sensitivity takes place. When mechanical packing of this nature occurs, the initial sensitivity can be restored only by tapping the transmitter until the granules assume their original configuration. By filling the chamber about 5 per cent less than full, it has been found possible to avoid this packing without effecting an appreciable increase in the initial resistance and noise or the aging rate.

Breathing

In many granular carbon transmitters, variations in resistance and sensitivity of considerable magnitude occur from changes in the dimensions of the various parts of the carbon chamber, caused by the heating effect of the direct current. These changes in resistance usually take place rather slowly and often extend over a period of several minutes. It is customary to refer to them as "breathing." In a handset transmitter, it is particularly important to eliminate any sources of variation of this kind, since they are likely to be more serious in their effects.

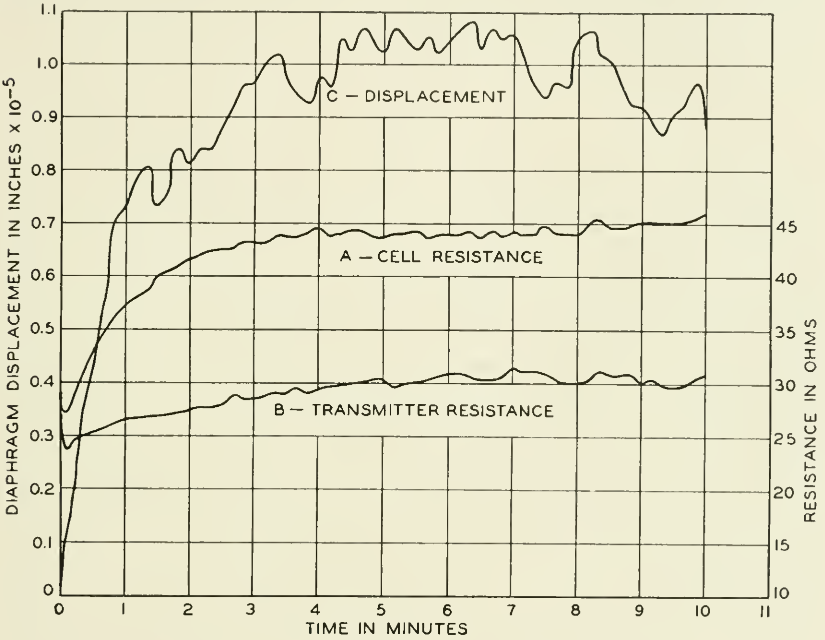


Fig. 9—Transmitter breathing.

Interesting means of breathing control have been provided. If the diaphragm is removed and a block of lavite substituted, the resistance after being connected in circuit changes with time in the manner shown by curve *A*, Fig. 9. The large increase in resistance is caused by the expansion of the walls of the chamber and a reduction in the contact pressure between the granules. An increase of this magnitude would be objectionable in commercial use, for it would not only make it more difficult to meet the resistance requirements but would also increase the contact voltages, especially in aged instruments, to a point where undesirable carbon noise would result. To counteract the effect of the expansion of the walls of the carbon chamber, the portion of the diaphragm in contact with the carbon is coned slightly, so that, as the walls expand, the diaphragm moves inward sufficiently to maintain substantially constant contact pressures between the granules. The extent to which this feature is effective in reducing breathing is shown by curve *B*, Fig. 9.

Cohering

In common with a number of other materials, granular carbon is susceptible to cohering. When cohering occurs the resistance and sensitivity of the carbon are lowered and remain so until the transmitter is subjected to mechanical agitation. Experience has shown that cohering will greatly reduce the output of the handset transmitter and that fairly loud talking or a sharp mechanical shock is required to restore it to its initial sensitivity. Not infrequently cohering results from breaking the circuit connecting the transmitter to the battery, as for example, when the subscriber depresses the switch in the mounting in order to attract the attention of the operator. A study of the electrical conditions responsible for cohering under these circumstances has shown that the distributed capacity and the inductance of the component parts of the station set are such that transient oscillations of a frequency of several thousand kilocycles per second are set up by the breaking of the circuit. Further investigation has shown that the transmitter can be protected from the cohering effect of these oscillations, without introducing a transmission loss at voice frequencies, by connecting a condenser of a few thousandths of a microfarad capacity across the transmitter to by-pass these transient currents.

In the case of the deskstand transmitter, these cohering effects have not been important for several reasons. Probably the principal one is that the mechanical impact incident to switchhook operation is carried directly to the transmitter and prevents appreciable cohering.

Carbon Preparation

Substantial improvements in the technique of granular carbon manufacture have been made in conjunction with the development of the new transmitter. Roasting processes have been adopted which subject each granule to substantially the same heat treatment and produce carbon of unusually uniform electrical properties. Uniformity is further insured by the use of a magnetic separator to remove granules having an undesirably high iron content, and the removal in an air stream of flat, wedge-shaped particles which tend to cause carbon leakage in the transmitter by working their way between the layers of paper which form the closure between the carbon chamber and diaphragm.

Measures of this kind have been of particular value, not only in bettering the transmission characteristics of the individual transmitter, but in securing a uniform commercial product. Without these improvements, indeed, it is doubtful if the production of the transmitter could be maintained at the high rate now required without a sacrifice in the average quality of the product.

Receiver

The parts of the handset receiver are assembled in a die cast aluminum housing, and form a unit which mounts in a threaded bushing in the handle in much the same manner as the transmitter unit. One connection to the winding is made through the threaded portion of the case, the other through a contact spring in the base. The cap and spacing ring are made of phenol plastic and thoroughly insulate the metal parts so that the user cannot come in contact with any portion of the receiver which forms a part of the electrical circuit. The spacing ring also serves as a lock ring for holding the receiver on the handle. A grid in the cap prevents damage to the diaphragm from the projecting portions of the mounting. The layout of the holes in the grid is such that a dent in the diaphragm caused by inserting a pencil or other sharp object through the grid, will not occur at a point over the pole faces and interfere with the operation of the receiver.

While in general the design of this receiver follows conventional lines, the choice of materials and the design of the magnetic circuit have resulted in an increase in efficiency of several db as compared with the deskstand receiver. It has not been considered desirable, however, to make use directly of this increase in efficiency for several reasons.

Although greater receiving efficiency raises the intensity of the incoming speech, it also increases the intensity of any noise present on the line or picked up as sidetone. Since, therefore, the ratio between the received speech and the noise is not improved, the general reception would be little benefited by an increase in receiving efficiency alone. Furthermore, the increased sidetone causes an effective loss in transmitting level by causing the speaker to lower the intensity of his speech.

The higher efficiency of the handset receiver is, however, of appreciable indirect value, and has been used in two ways, at the same time maintaining the same intensity of the received speech as is obtained with the deskstand.

By lowering the impedance to about half that of the deskstand receiver, the loss in the sidetone path and the transmitting efficiency of the set have been increased. The lower sidetone tends to increase the speaker's talking level, giving a further effective gain in transmitting efficiency. The reduction in the room noise appearing in the sidetone also results in an effective receiving gain as compared to the deskstand.

The separation between the diaphragm and the pole pieces has been increased so as to minimize the "freezing" of the diaphragm on the pole pieces, which is likely to occur when clamped diaphragm receivers are subjected to sudden changes in temperature of considerable magnitude.

Assembled Handset

The dimensions and proportioning of the handle and the assembled handset are the result of rather extensive investigation. They represent an effort both to meet the technical requirements and to produce a mechanically rugged structure, light in weight and comfortable to hold, as well as attractive in appearance and in harmony with the mounting.

Prevention of Howling

One of the major objections to the handset in the past has been the tendency to howl. While a certain degree of howling control can be realized in the electrical coupling by the design of the station circuit, such control cannot be depended upon to take care of extreme circuit conditions where the coupling is greatest and howling most likely to occur.

From an economic standpoint, moreover, it is desirable that the handset should be completely interchangeable with the deskstand without the necessity for changes in the set or other parts of the telephone plant. If this is to be realized, the handset must be so designed as to be inherently free from the danger of howling under any conditions encountered.

Mechanical disturbances set up by the receiver produce transverse vibration in the handle which moves the transmitter as a whole. The solid phenol plastic handle employed in the present handset has a relatively high resonant frequency and provides a comparatively inefficient medium for the transmission of mechanical disturbances from the receiver to the transmitter in the frequency range where the instruments respond most readily.

A further and perhaps a major factor in eliminating the danger of howling is to be found in the characteristics of the transmitter. The response of a transmitter to bodily movement of the whole structure is of course very different from its response to acoustical pressures, and depends largely on the mechanical impedance of the diaphragm and its supporting structure. The extremely light and highly damped diaphragm of this transmitter is ideally suited for minimum response to mechanical vibration.

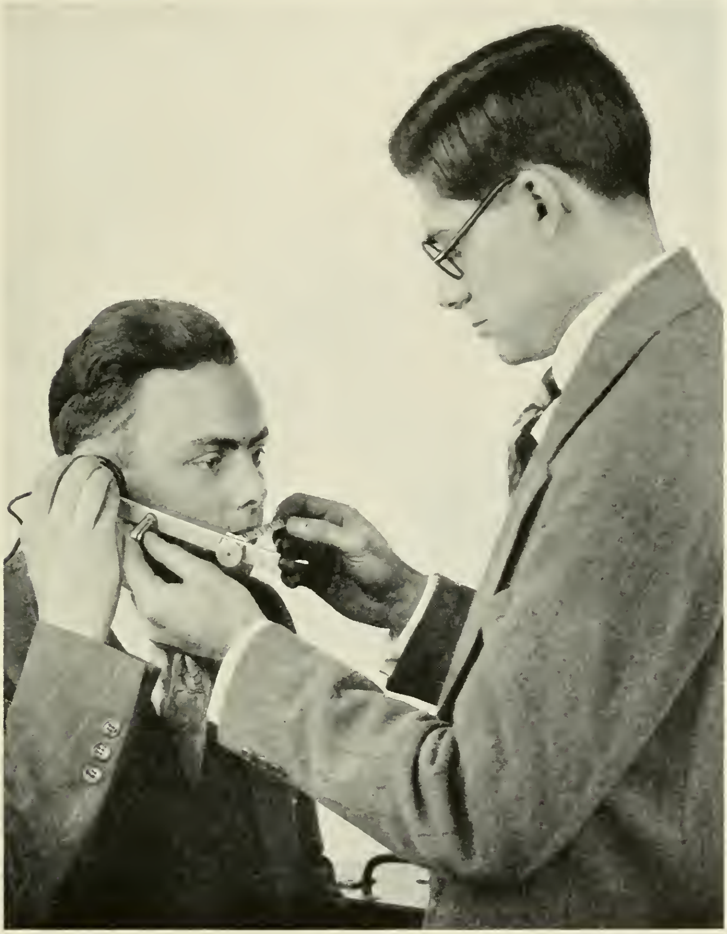


Fig. 10—Use of gauge for head measurements.

Under the worst conditions encountered, a margin of at least 15 db is realized with this structure which effectively eliminates howling as a limiting factor in the general use of the handset.

Determination of Dimensions

Particular effort has been made to proportion the handset so that the transmitter mouthpiece may be as near as possible to the lips of the talker, thereby avoiding unnecessary transmitting loss. Care has been taken also to provide adequate clearance between his hand and cheek. In determining the proper dimensions of the handle and assembly, about 4000 measurements of head dimensions were made. The subjects selected for these measurements included both sexes and the various races in about the proportion indicated by the census figures.

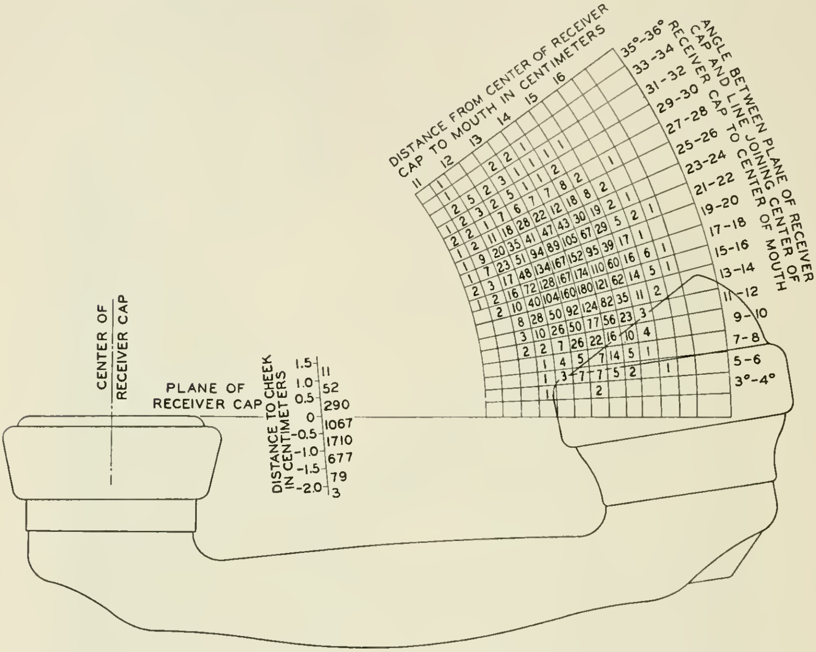


Fig. 11—Distribution of the head measurement data.

The head measurements were made with a special gauge as shown in Fig. 10. The receiver cap was held to the ear and the pin placed in contact with the cheek. The slider was moved along the calibrated bar until it touched the lips. A record was made of the dimensions which determined the position of the center of the lips relative to the center of the receiver cap and the clearance between the handset handle and the cheek.

The data obtained are shown graphically on Fig. 11 together with an outline of the handset drawn to the same scale. All except about 3 per cent of the persons shown on the chart can use the handset by holding the receiver to the ear in a normal manner. The others can be accommodated by a slight shift of the receiver on the ear. The effect of such a shift on received speech, for a small number of persons, is unimportant in comparison with the large improvement in transmitting performance for the great majority which results from the decrease in the average distance of the mouthpiece from the lips.

PERFORMANCE IN SERVICE

Since the initial introduction of the handset, close contact has been maintained with its performance by tests and observations under actual service conditions, and by the examination of instruments returned from service.

It has been shown by these observations that, although the output of the handset transmitter as used by the subscriber is on the average lower than that of deskstand transmitters of the most efficient types, the improved response and articulation are an adequate compensation for the lower level. The transmission performance of the handset as rated by the repetitions,[7] has been found to be as good as that obtained with the deskstand.

Undesirable variations in transmission and resistance with change in position, excessive carbon noise, and howling, all of which have heretofore presented serious obstacles to the adoption of a handset for general use, have been successfully overcome in the design which has been described. It has been found practicable to use this handset interchangeably with the deskstand in the existing telephone plant, without important reactions on either transmission or signalling performance.

That the design, in addition, meets the desires of the public for the convenience of a hand telephone set is best evidenced by the steady increase in demand to more than one million a year at the present time.

---

[7] "Rating the Transmission Performance of Telephone Circuits," W. H. Martin, *Bell System Technical Journal*, Jan. 1931, pp. 116–131.

Precision Methods Used in Constructing Electric Wave Filters for Carrier Systems

By G. R. HARRIS

Electric wave filters are used extensively in carrier telephone and telegraph systems. In order that such systems may be operated efficiently and economically, the requirements placed on the filters they employ are occasionally so severe in nature that new methods in design or construction must be developed to make the commercial production of the filters possible. The band filters for the Type "C" Carrier Telephone System are cases in point. This paper sets forth the requirements which were met in the design of the filters for this system, and describes a new manufacturing adjustment made necessary by these requirements. This feature consists essentially of an inductance continuously variable over a small range above and below its nominal value; the adjustment is not used to set the coil inductance at its specified value, but to locate correctly the series or parallel resonance of the mesh of which the coil is a part. The bridge and associated apparatus developed to facilitate this adjustment are also described.

ONE of the most important fields of usefulness of the electric wave filter in the Bell System is found in carrier current telephone and telegraph systems. A carrier telephone system transmits several messages over the same line by employing several carrier currents of frequencies higher than those in the ordinary voice band, and modulating these carriers with the messages to be transmitted. The messages are then transmitted over the line as side bands of their respective carriers, and are demodulated to their original voice frequencies at the receiving end.

Usually only one side band of a carrier is transmitted in order to reduce the frequency space required for each channel. Each message as it appears on the line occupies its own portion of the frequency spectrum, distinct from that occupied by any other message, and is transmitted without interference from other messages; but in order that the operations of modulation and demodulation may be carried on without interference and in order that unwanted side bands may be suppressed, it is necessary that each modulator and demodulator be equipped with some apparatus which will pass all the side band frequencies making up one message and reject all others. Electric wave filters are the instruments used for this purpose, since they possess the property of passing currents of certain chosen frequencies with very small loss, and of offering high attenuation to other chosen frequencies.

The theory of electric wave filters as used in carrier systems has been discussed in previous articles in this *Journal*, notably "Physical

Theory of the Electric Wave Filter" by G. A. Campbell, which appeared in the November, 1922 issue, and "Theory and Design of Uniform and Composite Electric Wave Filters" by O. J. Zobel, in the issue for January, 1923. It is not the purpose of the author to discuss further the theory of wave filter design or operation, but rather to discuss some of the problems that arise when an attempt is made to construct filters which must meet certain requirements of a carrier system with a very small margin of variation, and to describe the methods of solution adopted. It is assumed that the reader is familiar with the general principles of electric wave filter theory.

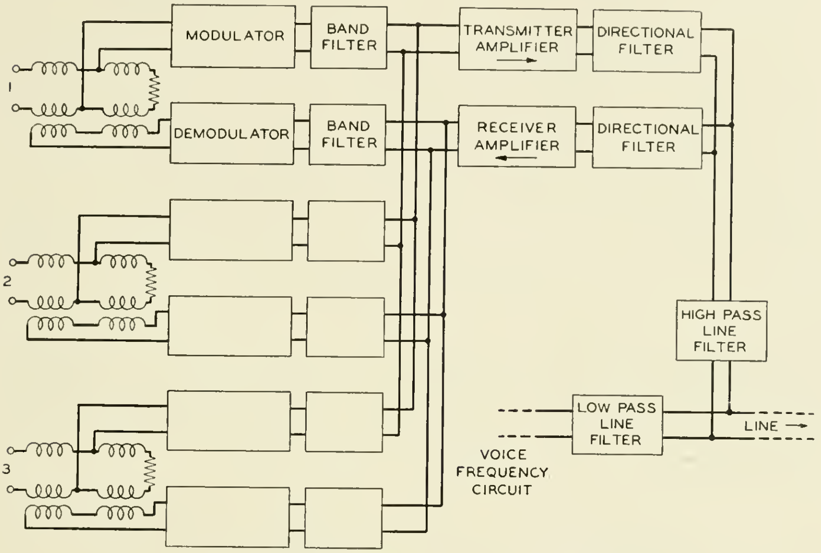


Fig. 1—Type "C" system terminal arrangement.

Different carrier systems require different numbers and designs of filters. A carrier system designed to provide three two-way telephone channels would have six band filters at each terminal. These filters would be of a type that passes all frequencies between certain upper and lower frequency limits, and provides a high attenuation or loss to all frequencies below and above these limits. A simplified diagram of such a terminal arrangement is shown in Fig. 1.

In addition to the band filters, a pair of directional filters is required at each terminal to separate the three incoming from the three outgoing channels. This directional filter pair is composed of a low-pass and a high-pass filter, which as their names imply, pass respectively all frequencies from zero up to the predetermined cut-off point and attenuate those above this point, or pass all frequencies above the cut-off point and attenuate all those below. The cut-off points of this pair of filters are so arranged that the passing region of the low-pass filter includes the three lower frequency channels and the passing range of the high-pass filter includes the three upper frequency channels. The location of the various filters in the frequency spectrum is shown in Fig. 2.

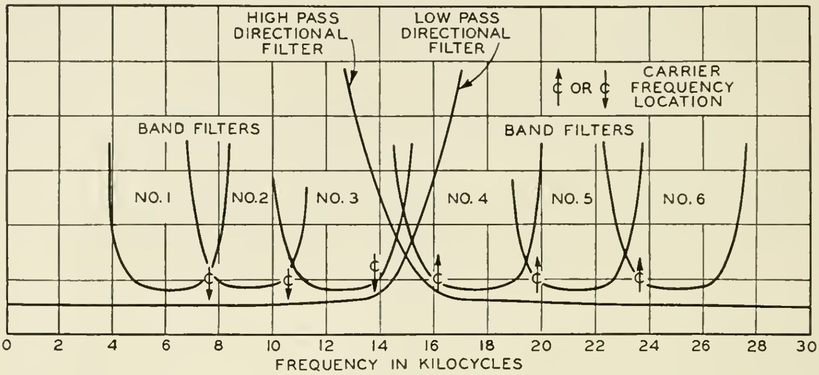


Fig. 2—Frequency spectrum of type "C" normal system.

If the carrier system is long enough to require repeaters, additional directional filters must be employed. These filters are required because repeaters must amplify frequencies going in both directions. The arrangement of the repeater is shown in Fig. 3.

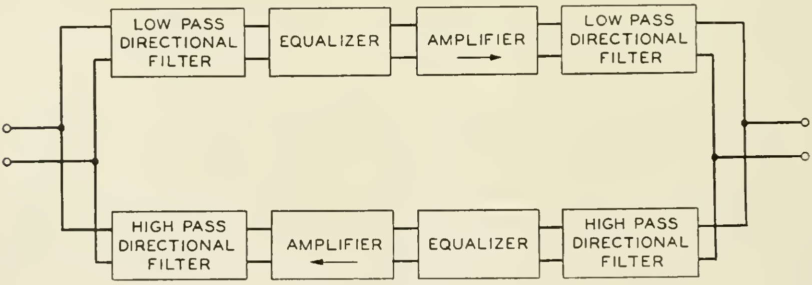

LOW PASS DIRECTIONAL FILTER
EQUALIZER
AMPLIFIER
LOW PASS DIRECTIONAL FILTER
HIGH PASS DIRECTIONAL FILTER
AMPLIFIER
EQUALIZER
HIGH PASS DIRECTIONAL FILTER

Fig. 3—Schematic of type "C" repeater.

This paper will be concerned with the filters for the Western Electric Type "C" Carrier Telephone System,[1] since the requirements placed on the apparatus for this system raised certain problems in filter design and construction which had not then been satisfactorily solved. The highest frequency to be used in this system was determined, of course, by the limitations of wire transmission as the frequency increases. The lowest was set chiefly by the desire to avoid interference from carrier telegraph systems on other wires of the same pole line. Between these limits, the six bands (three two-way channels) had to be located. At the same time it was desirable to provide a second system, with a frequency allocation offset or "staggered" with respect to the first, to permit operation of carrier systems on pairs in close proximity where the normal crosstalk would not permit operation of the same systems. This is shown graphically on Fig. 4. The crosstalk advantage thus obtained is rather critical with respect to the overlap between interfering channels. These two factors, first, interference between channels in the same system and, second, interference between channels in adjacent staggered systems, emphasized the need for a high degree of precision in the location of the individual bands.

---

[1] H. A. Affel, C. S. Demarest and C. W. Green, "Carrier Systems on Long Distance Telephone Lines," *Bell Sys. Tech. Jour.*, July, 1928.

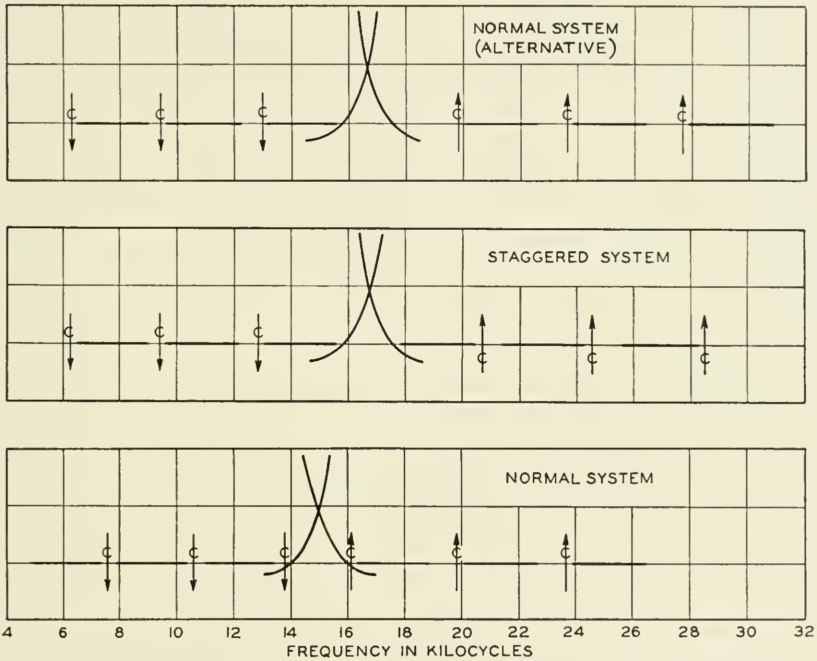


Fig. 4—Frequency spectra of type "C" normal and staggered systems.

The theoretical requirements of the filters, although severe, were met without any departure from conventional design practices. The calculated attenuation characteristics satisfied the minimum loss requirements and did not exceed the maximum loss requirements in the transmitting bands. The margins, however, were not very great.

It is evident, therefore, from Fig. 5, that a very small shift in the frequency location of the loss characteristic of a filter would throw it outside the required limits. The shift which would cause this to happen was, in fact, so small, ± 125 cycles, that the manufacture of these filters by methods currently in use resulted in enough rejections at the factory to warrant the development of a more precise method.

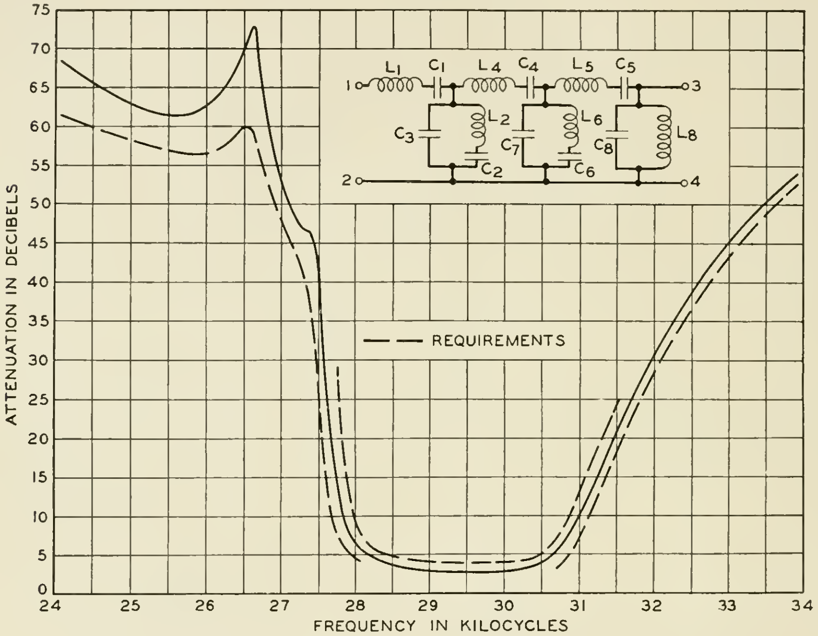


Fig. 5—Attenuation characteristic of 38–N filter, showing requirements.

The attenuation characteristic of a manufactured filter seldom conforms exactly with that desired because of manufacturing tolerances which must be allowed in its component elements. The inductances and capacitances of the coils and condensers which make up a filter differ somewhat from their specified values for several reasons. The precision of their adjustment can be no greater than the precision of the circuits in which they are measured. It is usually less than this figure because the coils and condensers used in precision filters are potted in a moisture-proofing compound, after adjustment, and this potting produces a small change in the capacitance of the condensers and the distributed capacitance of coils which is not uniform and cannot be allowed for exactly in advance. Coil adjustment is limited by another factor, as well; the adjustment of a single coil must be made in units of one turn. For these reasons it is very difficult to adjust condensers to better than about ± 0.3 per cent, and coils to better than a little over ± 1 per cent.

The effect of the variation in element values may be manifested in several ways. In a simple half-section low-pass filter, for example, the cut-off frequency

$$f_C = \frac{1}{2\pi\sqrt{LC}}$$

where L and C are the values of the series inductance and shunt capacitance. A variation in L or C will produce a change in the cut-off frequency proportional to the change in the element value. Changing the cut-off frequency is, of course, equivalent to shifting the whole attenuation characteristic upward or downward in frequency. A change in an element value also produces a change in the filter impedance. Suppose that in a network composed of a number of series and shunt arms, the impedance of one arm of the network, Z_q is increased by an amount ΔZ_q. The current through that branch is I_q. The change in impedance is approximately equivalent to introducing a voltage e in the circuit.

$$e = -I_q \Delta Z_q.$$

If E is assumed to be the voltage in series with the terminal impedance Z_s at the sending end of the network, and I_s is the input current,

$$-\frac{I_q \Delta Z_q}{\Delta I_s} = \frac{E}{I_q}$$

from the reciprocity theorem. But

$$\frac{\Delta I_s}{I_s} = -\frac{\Delta Z}{Z + Z_s},$$

where Z is the input impedance of the network, and ΔZ is the change in Z produced by the impedance change in the arm Z_q. Furthermore,

$$E = (Z + Z_s) I_s.$$

Therefore

$$E = -I_s^2 \frac{\Delta Z}{\Delta I_s}.$$

And

$$\Delta Z = \Delta Z_q \left(\frac{I_q}{I_s}\right)^2.$$

As a filter is usually designed to match two fixed impedances between which it must operate, a change in its impedance caused by element variation will produce a mismatch between the filter and its terminating impedances which will result in reflection loss in the transmitting region. However, the reflection loss arising from impedance deviation caused by element variation is usually much less troublesome than the shift in the attenuation characteristic produced by this variation.

A more complicated type of filter, such as the symmetrical "Constant k" band pass section exhibits a more involved relation between the cut-off frequencies and the values of the elements. L_1 and C_1 are the series arm inductance and capacitance, while L_2 and C_2 are the inductance and capacitance in the shunt arm.

$$\left.\begin{matrix} f_2 \\ f_1 \end{matrix}\right\} = \frac{1}{2\pi}\left[\sqrt{\frac{1}{L_1C_2}+\frac{1}{L_1C_1}} \pm \sqrt{\frac{1}{L_1C_2}}\right].$$

In the equation as written L_2 does not appear explicitly, but as in this type of section $L_1C_1 = L_2C_2$ the equation may be written with L_2C_2 as the denominator of the second term in place of L_1C_1, if desired. Therefore, three LC products influence each cut-off frequency; three resonant frequencies, therefore, must be held constant if the cut-offs are not to vary. The mid-band frequency of this section is given by the equation

$$f_m = \frac{1}{2\pi}\sqrt{\frac{1}{2}\left[\frac{1}{L_1C_1}+\frac{1}{L_2C_2}\right]},$$

from which it is seen that two LC products, L_1C_1 and L_2C_2, must be held constant if f_m is not to vary.

Element variation may cause a shift in the cut-off points and thus a displacement of the attenuation or impedance characteristic of any type of section, but it may have another effect on the characteristics of M type sections. An M type section provides a peak of attenuation at some finite frequency, which is determined, in most sections of this type in common use, by a single pair of elements; either a series resonant combination shunted across the filter or a parallel resonant combination in a series arm. Variation in either of the elements determining the location of the peak will cause a shift in the location of this peak proportional to the element variation. Changes in these elements would affect the cut-offs of the section as well as the peak or peaks, but it may happen that variation in the opposite direction of other elements in the section will result in negligible shift of the cut-offs, while the peaks are shifted noticeably. Thus the attenuation characteristic of an M type section may be distorted as well as displaced by element variation.

The critical region of the type "C" system band filter requirements was the slope of the attenuation characteristic from the cut-off to the peak nearest the carrier side of the transmitting band. The system was designed to transmit one sideband only, and the band filters were required to suppress the carriers at least 15 db. The transmitted sideband for each filter extended from 250 to 2750 cycles from the carrier, so that in every case the filter had to offer at least 15 db suppression to the carrier frequency, which was only 250 cycles away from the edge of the transmitting band. Where the carrier frequency was 25 kc., the separation was only 1 per cent. This requirement necessitated very steep slopes of the attenuation characteristic, especially for the higher frequency filters; at these slopes a frequency shift of 1/2 per cent produced a change in attenuation of 9 or 10 db, as may be observed in Fig. 6.

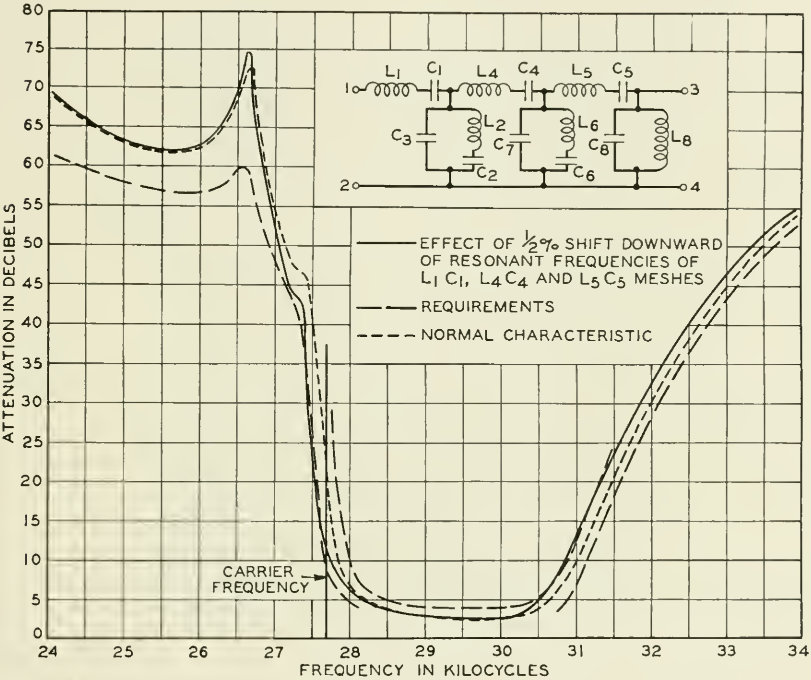


Fig. 6—Attenuation characteristics of 38–N filter, showing effect of element variation.

As the precision of adjustment for condensers was three or four times that for coils, it was obvious that some improvement in coil construction or adjustment would have to be made. Furthermore, since the important limiting factors in the adjustment of both coils and condensers were the precision of the bridge standards and the bridge operating technique in the shop it seemed unlikely that more precise location of the filter characteristics could be obtained until these factors were improved. However, another way was found. It has been pointed out in preceding paragraphs that critical points, such as cut-off points and attenuation "peaks" are usually dependent upon the resonant frequency of a pair of elements, or perhaps of several pairs. It is much more important that these critical frequencies be correctly located; that is, that the L–C product of the elements be of the correct value, than that either element alone be of the correct value. Therefore, if some way could be found of adjusting a combination of elements associated together in a filter so that the L–C product would be of the correct value, a considerable improvement in the precision of the attenuation characteristic location might be expected. If the elements could be adjusted so that a negative deviation in condenser capacity, for example, could be compensated for by a positive deviation in coil inductance, the L–C product, and therefore the resonant frequency, would be correct, and the effect of the individual element deviation would be greatly minimized, though not entirely eliminated. Either an adjustable coil or an adjustable condenser would make this procedure possible; and, as such a resonant frequency adjustment could be made to the precision of an oscillator calibration, or about 0.05 per cent, if necessary, the errors of inductance and capacity bridge standards, as well as initial adjustment error and variation due to potting would be eliminated at the resonant frequency, and their effects minimized greatly at other frequencies.

Development was therefore started on an adjustable type of inductance. As the simplest and most inexpensive coil was, of course, the most desirable, efforts were made to modify the solenoidal air core coil used in previous carrier filters. One scheme considered was to mount a small copper vane in such a way that it could be moved into or out of the field of the coil; the eddy-currents in the vane set up a small counter field that neutralized a part of the field of the main coil and thus reduced its inductance. Another method made use of a small "pancake" wound coil mounted vertically on the side of the main coil so that it could be moved up or down, into or out of the field of the main coil. This small coil was short-circuited, and therefore produced an effect similar to that of a copper vane. The model finally adopted, however, although using the same principle of adjustment, was made in a somewhat different form. A rectangular rotor was constructed to fit around the main coil. This rotor was pivoted to the main coil on its short sides and could be rotated through an angle of approximately 60° about its horizontal axis. On this rotor was wound a large number of turns of small gauge wire, and the winding was short-circuited. A coil of this type is shown in Fig. 7. The coupling between the rotor in its horizontal position and the main coil is very small, and therefore the effective inductance between terminals is practically that of the main coil. However, if the rotor is rotated in such a way that its plane becomes more nearly parallel with that of the main coil the coupling between it and the main coil is increased, and the effective inductance of the whole is reduced. It is possible to vary continuously the inductance of a coil of this design over a range of about $3\frac{1}{2}$ per cent.

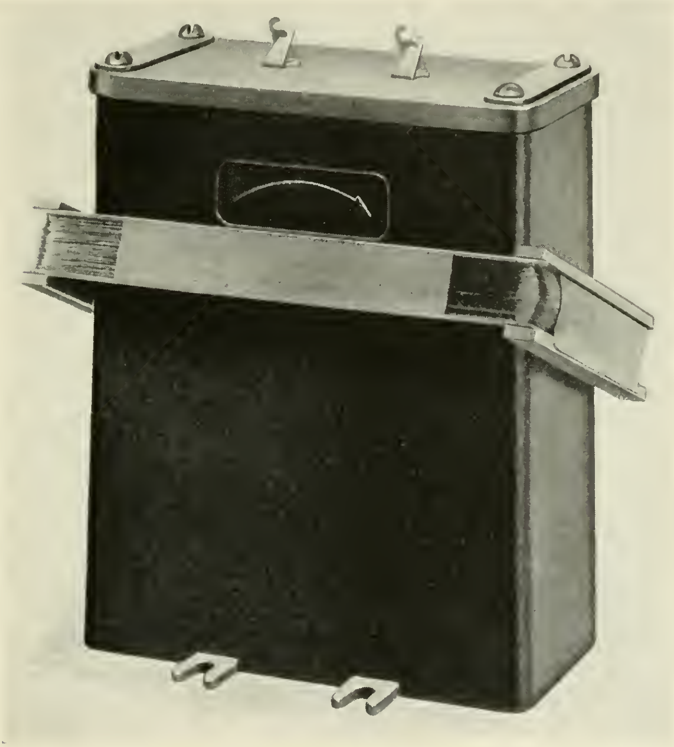


Fig. 7—Photograph of "AA" coil.

When coils of this type and their associated condensers are connected together as they would appear in a filter, their resonant frequency or frequencies are located by an adjustment of the coil or coils in the mesh. If the mesh contains more than one adjustable coil, more than one resonant frequency may be located. As many of these resonant frequencies may be located in this way as there are adjustable elements in the mesh. Since the condensers are not adjustable, all of the frequencies of series and parallel resonance in a complex mesh, will not, in general, be capable of adjustment. If the mesh is composed of n condensers and m coils, and the difference between m and n is not more than 1 (if this condition is violated the mesh can be reduced to an electrically equivalent one containing a smaller number of elements), the number of resonant and anti-resonant frequencies is clearly $(m + n - 1)$. However, since the number of adjustments possible is equal only to the number of coils, m, the number of frequencies which cannot be exactly located is $(n - 1)$, or one less than the number of condensers. This condition is not very objectionable in practice as the mesh configuration is usually chosen to be of a form which permits the more important critical frequencies to be adjusted. In a four-element mesh, the largest usually encountered in an ordinary filter structure, two of the three resonant frequencies may be adjusted, and furthermore, the adjusted frequencies may be any two of the three; in most cases the adjustment of two frequencies is sufficient.

The frequency adjustment is actually made on a bridge circuit, the schematic of which is shown in Fig. 8. Two of the bridge arms are, of course, ratio arms; the mesh to be adjusted is connected in circuit as a third arm, and the fourth arm is a variable resistance. For series resonance adjustment, the desired frequency is impressed on the bridge, and the inductance of the adjustable coils varied so that the impedance of the mesh becomes a pure resistance, which is balanced by an adjustment of the resistance in the opposite bridge arm. For a parallel resonance measurement, a resistance of 10,000 ohms is shunted across the mesh being adjusted, and the coil varied until the reactance component of the mesh impedance vanishes, when the resistance of the 10,000 ohms in parallel with the resistance of the mesh can be balanced in the fourth bridge arm.

This frequency adjustment of filter arms involved a change in filter manufacturing procedure. Ordinarily the coils and condensers for the filter are individually adjusted, then mounted, wired, and the filter finally tested as a whole, but an additional step is required in the construction of filters employing adjustable coils. It is necessary to adjust the resonant frequencies of meshes after the elements are mounted, but before the filter is finally wired and tested, and furthermore, it is necessary to adjust these elements under exactly the same conditions that obtain during filter operation if the full benefit of the adjustable feature is to be realized. This requirement presented a difficulty, as the filters for the type "C" carrier system are assembled in copper boxes or shields as shown in Fig. 9 and are completely soldered all around, including the lid. This type of assembly was made necessary by the requirements on cross-talk between filters mounted next to each other on a vertical rack. The cross-talk between two filters mounted side by side was required to be 135 db below the signal currents, or, in other words, the ratio of the desired signal to the cross-talk had to be a little over 5,000,000 : 1. The chief cause of such cross-talk between two pieces of apparatus is electromagnetic coupling between them; to reduce this coupling eddy-current shielding is employed. Eddy-current shielding is accomplished by surrounding the source of a varying electromagnetic field—a coil, say, or the coils in a filter—by some metal, such as copper, brass or aluminum, which has good conductivity.

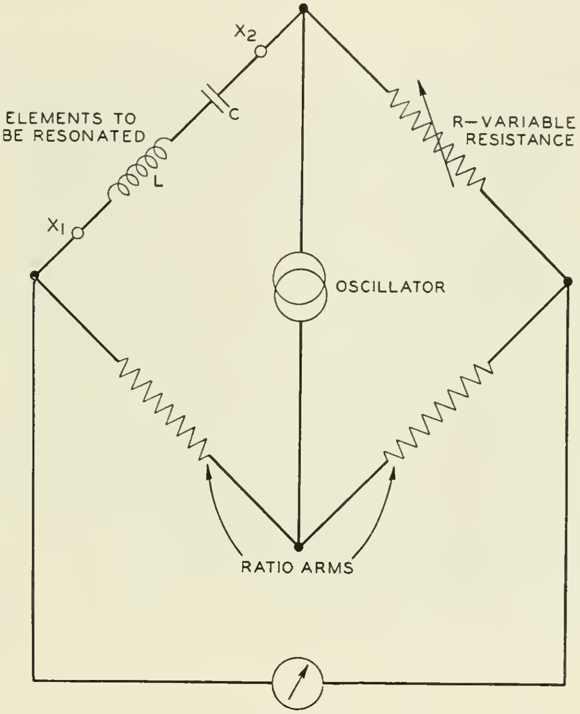


Fig. 8—Simplified schematic of L–C bridge.

To make clear the effect of a shield on an inclosed inductance, a brief discussion of shielding may be of interest. If a coil through which an alternating current is flowing is enclosed in such a shield, one part of the flux lines from the coil will encounter the shield; the remaining part will not. Let every line $\Delta\varphi$ of the flux which encounters the shield be divided into components perpendicular to the shield and parallel to it. If Θ is the angle between such a line and the plane of the shield, the component perpendicular to the shield, which is effective in setting up eddy currents in it is $\Delta\varphi \sin \Theta$. The other component, $\Delta\varphi \cos \Theta$, makes no coupling with this shield, and sets up no eddy currents. To every line $\Delta\varphi \sin \Theta$ the shield presents a certain inductance and resistance, and the average inductance and resistance presented to all the linking lines may be taken as the inductance and resistance of the shield. The equivalent circuit of a coil surrounded by a shield may be indicated, then, as a coil L_1 of two sections coupled to a second coil L_2, in parallel with a resistance R_2. This parallel circuit is the equivalent circuit of the shield. The coil L_1 has a total inductance equal to that of the coil in the shield, and is considered to be divided into two sections, the first of which, "A," makes no coupling at all with L_2, and the second, "B," which makes perfect coupling with L_2. If K is the coefficient of coupling between the coil L_1 as a whole and L_2,

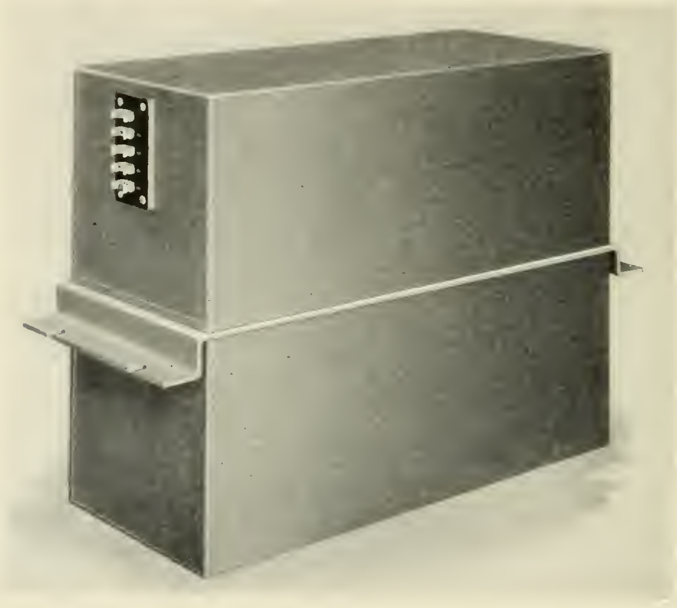


Fig. 9—Photograph of filter in soldered shield.

the portion "B" of L_1 is equal to K^2L_1. Now let a current I_1 be assumed to be flowing in L_1 and, I_2 in L_2. The mutual inductance between the coil and the shield is $-K\sqrt{L_1L_2}$. If e_1 is the voltage drop across the portion "B" of L_1

$$\begin{cases} e_1 = j\omega K^2L_1I_1 - j\omega K\sqrt{L_1L_2}I_2 \\ 0 = j\omega L_2I_2 + R_2I_2 - j\omega K\sqrt{L_1L_2}I_1, \end{cases} \tag{1}$$

from which

$$e_1 = \frac{I_1K^2R_2L_1\omega}{R_2^2 + L_2^2\omega^2}(L_2\omega + jR_2).$$

Suppose $R_2 = 0$, which would be the case if the shield were made of a perfect conductor. Then $e_1 = 0$. If, however, R_2 is assumed to be infinite, $e_1 = I_1j\omega K^2L_1$, the voltage which would appear if the shield were not present. Since the portion "B" of L_1 is the only inductance coupled with the shield, it is the only source of a field which might extend beyond the shield. Therefore, when R_2 is 0, and e_1 the voltage across "B" is also 0, there is no energy in the field, and it obviously cannot make coupling with anything beyond the shield. This condition would obtain for perfect shielding, but R_2, of course, is never zero in practice.

When R_2 is finite, "e_1," the voltage across "B" is finite, and a certain amount of energy is transferred to the equivalent circuit of the shield. Part of this energy is dissipated in the equivalent resistance of the shield, and the remainder is stored in the field of the equivalent inductance.

The reactance, or "j" term of Equation 1 is a measure of the energy stored in the shield and is, therefore, also a measure of the field which may exist outside the shield. As the flux is proportional to the ampere-turns, and the reactance is proportional to the square of the turns, the square root of the ratio of this reactance value when R_2 is infinite to its value when R_2 is finite is the average ratio of the flux φ_0 with no shield present to the flux φ_s with the shield present.

$$\frac{\varphi_0}{\varphi_s} = \sqrt{\frac{\frac{1}{R_2^2}}{R_2^2 + L_2^2\omega^2}} = \sqrt{1 + \frac{\omega^2}{P^2}}, \tag{2}$$

where $P = \frac{R_2}{L_2}$.

The efficacy of the shielding in reducing the external field is therefore seen to be dependent upon the frequency, and a parameter "P," the ratio of resistance to inductance of the shield itself. As the voltage induced in some coil beyond the shield is proportional to the flux, the reduction in cross-talk caused by shielding may be written

$$a = 20 \log_{10} \frac{\varphi_0}{\varphi_s} = 10 \log_{10} \left(1 + \frac{\omega^2}{P^2} \right) \cdot \tag{3}$$

In the 1/32 inch copper shields ordinarily used for filters, the parameter "P" has been found to have values from 1250 at 500 cycles to 2500 at 15000 cycles.

From Equation 1, the change in impedance, ΔZ of the coil, caused by the presence of the shield, may be written

$$\Delta Z = + \frac{[\omega^2 K^2 L_1 L_2 R_2]}{\omega^2 L_2^2 + R_2^2} - j\omega \frac{[\omega^2 K^2 L_1 L_2^2]}{\omega^2 L_2^2 + R_2^2} \cdot \tag{4}$$

This change is made up of a decrease in reactance and an increase in the effective resistance of the coil. The percentage change in both resistance and reactance may be written

$$\frac{\Delta R_1}{L_1} = - \frac{\omega^2 K^2 P}{P^2 + \omega^2}, \tag{5}$$

$$\frac{\Delta L_1}{L_1} = - \frac{\omega^2 K^2}{P^2 + \omega^2}, \tag{6}$$

and

$$\frac{\Delta R_1}{\Delta L_1} = - P. \tag{7}$$

The percentage change in the inductance of the solenoidal air-core coils mounted inside ranged from 2 per cent to 4 per cent, depending upon the proximity of the coil to the shield. Since these changes depended upon the distance of the coils from the shield, slight variations in the winding diameters of the coils caused variation in the magnitude of this cover effect, so that although this change in inductance was taken into account by specifying inductance values for the filter coils 2 to 4 per cent higher than indicated by the design computations, the corrections were only approximate. No such correction was, of course, necessary for the condensers; the condensers, therefore, as measured outside the filter shield, were still correct, but the coils under the same conditions, measured higher than specified values. It was obvious that the resonant adjustment should not be made with the coils outside their shields, as the approximate nature of the cover correction would vitiate the precision of the adjustment, but it was equally obvious that it would be very difficult to make a coil adjustment inside the soldered cans, as the elements would then be inaccessible.

As mentioned above, the purpose of the filter shielding was to reduce the cross-talk between filters mounted side by side; the effect of the shield upon the coil inductances was an unwelcome incidental. When the coil adjustment was considered, however, the shielding effect of the copper can was of no interest, but its effect on inductances was of paramount importance. Efforts were made to devise a dummy shield which would simulate the standard filter shield in its effect upon the inductances of the coils mounted inside but in which the coils would be readily accessible for adjustment. As it was expected that the adjustable coil would enable resonant frequencies to be located at their desired values within limits of ± 0.1 per cent, it was necessary that any dummy shield developed for use in the adjusting process should reproduce on the coils the effect of the filter shield to at least this figure. In the ideal case, it would have been desirable that the difference between the dummy or adjusting shield and the standard shield be no greater than the difference between two standard shields.

Several different designs of adjusting shields were tried in the attempt to realize this ideal. The model finally adopted was constructed in the following manner: A box of the same height and width but slightly shorter than the standard shield was built, open at one end. Laboratory tests showed that for coils placed inside the shield near the closed end, the reduction in inductance caused by the shield was almost entirely unaffected by conditions at the other end of the shield. The far end could be closed or left open with an effect which was negligible in comparison with the total inductance change caused by the shield. This condition, of course, held only for coils in the closed end or at a distance of at least three coil diameters from the open end of the shield. Since the maximum number of solenoidal coils mounted on a panel in the type "C" system filters was four, it was necessary only that the adjusting shield simulate the standard for two coil positions, as a filter panel could be turned end for end and the other two coils inserted in the shield after the first two coils were adjusted. The method of adjustment employed was the same for all the preliminary models as for the final design. Small holes were drilled in the top of the shield corresponding to the standard coil locations in the filters, and through these holes the coils were adjusted by means of hard rubber rods with small buttons at their ends. For each coil position a pair of these rods was provided, so arranged that one could be rested on each of the long sides of the coil cradles. Then, by pushing on one rod or the other, the operator could move the cradle to the desired position.

The adjusting shield as finally constructed, and as now used both by the Bell Laboratories and by the Western Electric Company, consists of two of the open shields previously described, mounted on an angle iron frame with the open ends facing each other, as illustrated in Fig. 10. The two shields, which are separated by a distance great enough to allow the filter assembly to be inserted between them, are equipped with a number of adjusting tools sufficient to adjust coils in any of the standard positions. The right hand shield was arranged to accommodate only the coils mounted on the right of the sub-panel center line, and the left hand shield was arranged to care for the coils mounted on the other half of the panel. Resonant frequency adjustments made in the shield have been checked to within a maximum of about ± 0.05 per cent against the frequency of the same combination in a standard soldered shield.

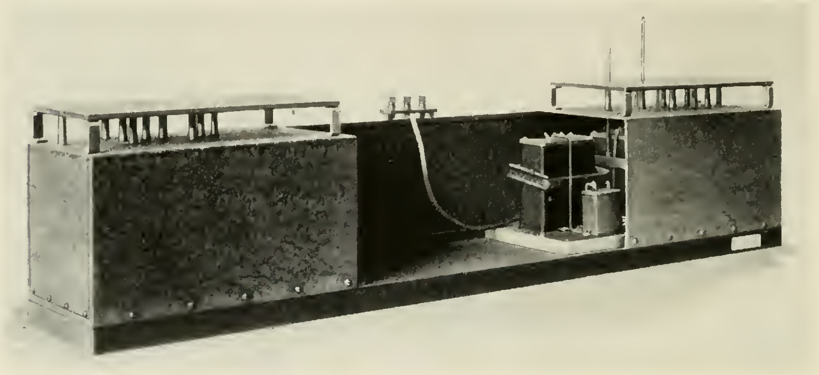


Fig. 10—Photograph of apparatus adjusting shield.

The shield is used with a special bridge and associated oscillator and detector, designed for the purpose. The bridge is very similar to the standard current bridges employed for measuring impedances at carrier frequencies, and the complete circuit is shown in Fig. 11.

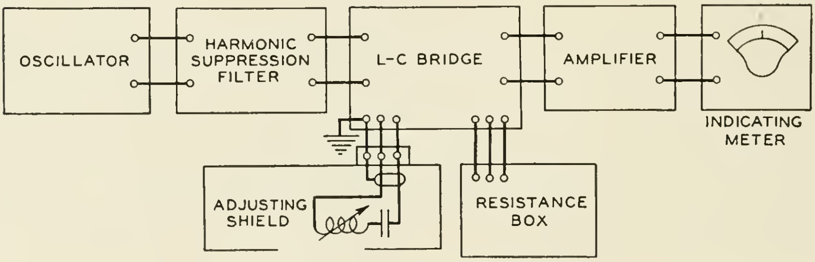


Fig. 11—L-C circuit arrangement.

A complete circuit for resonant frequency adjustment is set up at the Kearny, N. J. Works of the Western Electric Company where carrier filters are manufactured. There, after the coils and condensers in the filter are mounted upon their sub-panel (or panels), but before they are wired, the assemblies are brought over to the resonant frequency adjusting circuit, familiarly known as the *L–C* circuit. The sub-panel is placed in one end of the adjusting shield and the coils on that end of the sub-panel are adjusted in the manner previously outlined, after which the sub-panel is brought over to the other end of the shield, and the remaining coils are adjusted. After the coil rotors have been set at the positions for correct adjustment, they are locked by tightening the four screws which hold the coil terminal strip in place. The coil top presses down on the wedges (one of which is visible in Fig. 7) on each side of the coil case, and these wedges in turn clamp the rotor bearings firmly. Then the sub-panels are removed from the circuit, wired, and assembled in their shields.

This method of obtaining increased precision in filter manufacture was first employed in the band filters and equalizers for the Type "C" Carrier Telephone System. Previously, the ± 1 per cent limit on solenoidal air-core coils, the ± 0.3 per cent limit on mica condensers, and the additional ± 0.7 per cent variation in the effect of the filter shield in the coil inductance, caused by variation in the coil diameter, variation in the shield dimensions, and variation in mounting location, which added up to ± 2 per cent *L–C* variation, imposed a ± 1 per cent limit on the precision of frequency location. When the *L–C* method of adjustment is employed, the resonant frequency of a coil and condenser combination may be adjusted to ± 0.05 per cent, the precision of oscillator calibration, plus the ± 0.1 per cent limit on the adjustment process, or ± 0.15 per cent. In the laboratory it is possible to adjust the resonant frequency of an element combination to limits closer than ± 0.1 per cent, but the ± 0.1 per cent limits were set in the shop in order that production might proceed without undue delay occasioned by the process of adjustment. The ± 0.7 per cent tolerance caused by variation in coil diameter, shield dimensions and coil location is reduced to a variation of ± 0.1 per cent in *L–C* product, or ± 0.05 per cent in frequency; the difference between the adjusting shield and a filter shield caused by the variation in dimensions of filter shields. The precision of manufacture of filters using the *L–C* method of adjustment is, therefore, ± 0.15 per cent, the limits for the adjusting process, plus ± 0.05 per cent, the margin for variation in filter shield dimensions, or ± 0.20 per cent in frequency.

At 30 kc., this figure represents a variation in the location of the attenuation characteristic of ± 60 cycles. However, variations in the effective resistances of the filter coils, and variations in the reflection loss caused by small changes in the filter impedance produce changes in the attenuation of a filter which must be taken into consideration. The sum of these variations amounted to about ± 15 cycles. The ± 60 cycle precision at 30 kc., the highest frequency of the type "C" system, became, therefore, ± 75 cycles; a figure well within the ± 125 cycle maximum shop limit required by the system, and considerably less than the ± 300 cycle limits formerly obtainable.

The adoption of this method of filter construction has made it possible to realize in commercial manufacture the high precision of characteristic location which was formerly limited to laboratory construction.

Cathode Sputtering—A Commercial Application*

By HAL F. FRUTH

The theory of cathode sputtering with the advantages and limitations in the application of this process is discussed, followed by a description of the commercial equipment and methods for applying gold electrode surfaces to diaphragms of certain types of microphones.

By proper design of the vacuum chamber and the inside parts, a fairly uniform discharge current density and a uniform deposit is obtained. A constant sputtering rate is produced by the use of a bleeder valve which maintains a proper residual pressure. Adherence and continuity are obtained by the use of a special cleaning process.

An extensive bibliography on cathode sputtering is included.

INTRODUCTION

ALTHOUGH the process of electrostatic deposition of metals by high voltages in a partial vacuum, commonly known as cathode sputtering, has been known for more than a half century, it has heretofore found but little commercial application. Rather extensive use of it has, however, been made in physics research laboratories for such purposes as the production of highly reflecting surfaces on mirrors and prisms, for spectrometers, interferometers, etc., and the making of extremely thin metal films for fundamental studies in atomic structure and electron theory. Sputtering has also been used in the manufacture of very fine conducting quartz fibres for suspensions in sensitive instruments such as quadrant and string electrometers, galvanometers, and electrocardiographs, and, to some extent, for etching certain metals.

In the following paragraphs it is intended to give a brief explanation of the process and a description of a commercial application in the production of diaphragms for certain microphones.

THEORY

Some fifty years ago, various investigators working on high voltage discharges in vacuo discovered that disintegration of the cathode occurs for nearly all metals and that the removed metal is deposited in a very fine state of subdivision on nearby objects. Various theories have been formulated as to the mechanism of this phenomenon. Some investigators have attempted to explain it by stating that it is a type of electrical evaporation where the electrical potential is analogous to the temperature potential in ordinary thermal evaporation. More plausible theories, recently advanced, are that the metal atoms or particles leave the cathode as a result of positive ion bombardment, either as a result of single impacts or by cumulative action. This does not seem to hold for all metals and residual gases, and it has been suggested that in some cases sputtering is due to the absorption of radiation produced when the gas ions are stopped at the cathode.

---

* *Physics*, April, 1932.

There has also been a great deal of controversy as to the structure of the sputtered film. Most of the more recent evidence indicates that the films are of a crystalline nature.

Advantages of Sputtering Over Other Methods of Metallizing

In general, cathode sputtering cannot compete with the more common processes of metallic deposition, but in many special cases it can be employed where other methods are inapplicable. This is especially true (*a*) when metals are to be deposited upon non-conductors, (*b*) when the surfaces to be metallized would be injured by contact with chemical solutions or high temperature, (*c*) when either a very thin continuous metal coat or a very smooth, highly reflecting coat is desired, (*d*) when metals are to be deposited that are very difficult to deposit in any other manner, such as silicon, tellurium, or selenium, (*e*) when a metal is to be deposited upon another metal far removed from it in the electrochemical series, such as gold or platinum upon aluminum or magnesium. In such cases, electrolytic corrosion is very apt to occur in the presence of moisture or traces of electrolyte.

Commercial Application

Cathode sputtering has been used commercially by the Western Electric Company with excellent results in the manufacture of diaphragms for carbon broadcasting transmitters, illustrated in Fig. 1. This type of microphone consists essentially of a tightly stretched duralumin diaphragm located between two chambers containing granular carbon. The manner in which the double button type of microphone is connected to the amplifiers of the broadcasting, public address, or sound picture system is schematically shown in Fig. 2. Those parts of the diaphragm which come in contact with the carbon are covered with gold, a metal which has been found to be ideally suited for microphonic purposes. It is exceedingly important that the gold coating be continuous and remain adherent, for if the carbon should make contact in a number of places with the duralumin, it is quite likely that the microphone would be noisy and therefore unfit for use.

When the gold spots were formed by the plating process previously used, considerable trouble was experienced because of such defects as blisters, peeling, and pin holes probably due largely to electrolytic corrosion. The substitution of the sputtering process has practically eliminated such defects. Some additional advantage over electrolytic plating is that a thinner continuous film can be deposited, making it possible to stretch the diaphragm to a higher natural frequency with less tension. This decrease in tension has resulted in a smaller frequency loss due to fatigue. The introduction of the sputtering process has, therefore, not only improved the quality of the microphone, but lengthened its life.

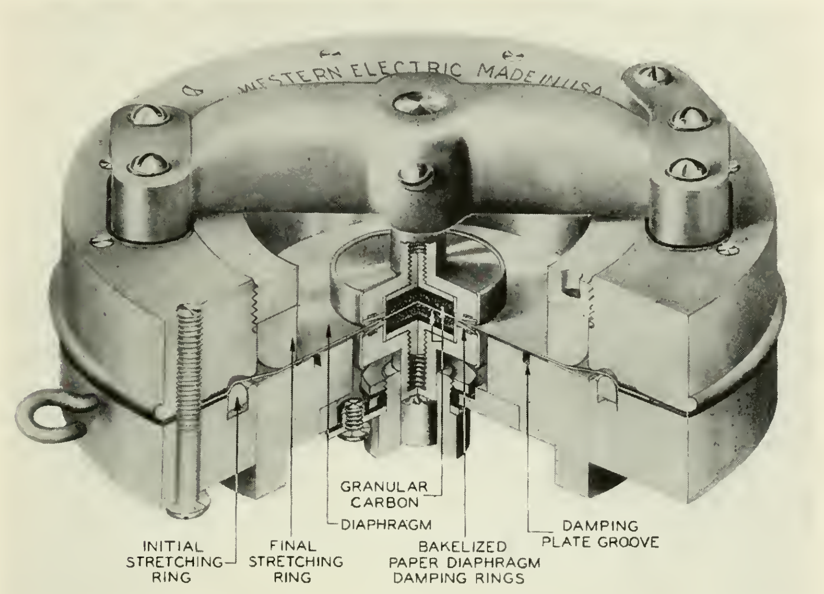


Fig. 1—Cross-sectional view of the two-carbon-chamber, stretched-diaphragm microphone.

Equipment and Method

In order to apply a satisfactory gold electrode surface to the duralumin diaphragm by cathode sputtering commercially, it was necessary to develop a special cleaning process for the diaphragms and suitable equipment with multiple sputtering electrodes.

To insure proper continuity and adherence of the gold to the duralumin, the glossy roll finish on the duralumin is broken up by means of a brass wire scratch brush to give a matte surface. The diaphragms are then scrubbed in acetone, rinsed in ether, and rubbed dry with filter paper to remove all traces of oil or grease. Great care is then taken to keep the spot to be coated free from all contaminations before sputtering.

The unit developed for commercial sputtering is shown in operation in Fig. 3. A two-gallon bell jar fitted upon a heavy ground pyrex glass plate contains six glass-covered aluminum cathodes, the exposed parts of which are fitted with renewable gold discs (Fig. 4). An adjustable aluminum stand to hold the diaphragm fixture and shield is placed upon the plate directly below the cathodes. The shield covers

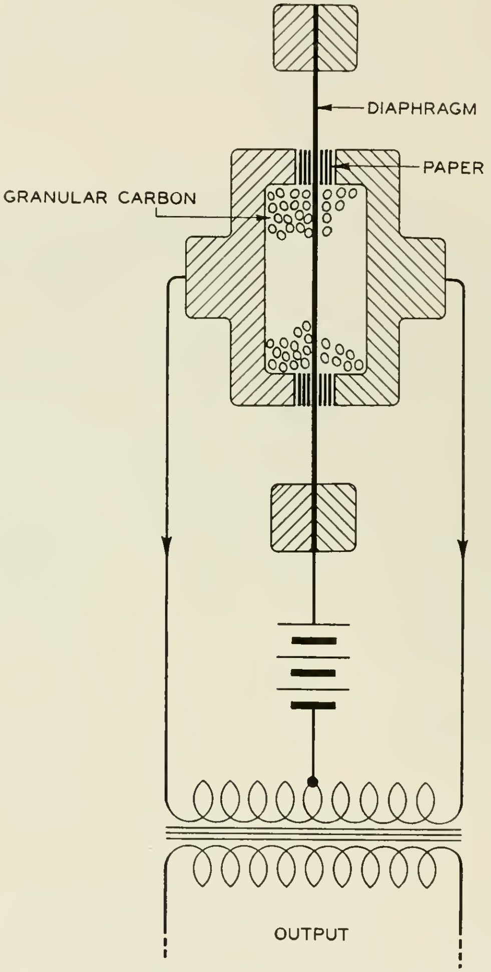


Fig. 2—Schematic diagram of conventional electrical connections of the microphone in a push-pull circuit.

all of each diaphragm except a central circular spot about 3/4″ in diameter. A vacuum is produced by means of a Megavac pump connected to the bell jar by 5/8″ pyrex glass tubing, which is made as short as possible to facilitate rapid pumping. In the vacuum system are a small rotating McLeod gauge, a trap, a bleeder valve, and an anode. The anode consists of a 20-turn, 1/2″ helix of No. 14 aluminum wire sealed into a one-inch tube at some distance from the bell jar. The end of the helix is shielded against high current density by a flanged piece of glass tubing. This type of anode has a life of at least two thousand hours.

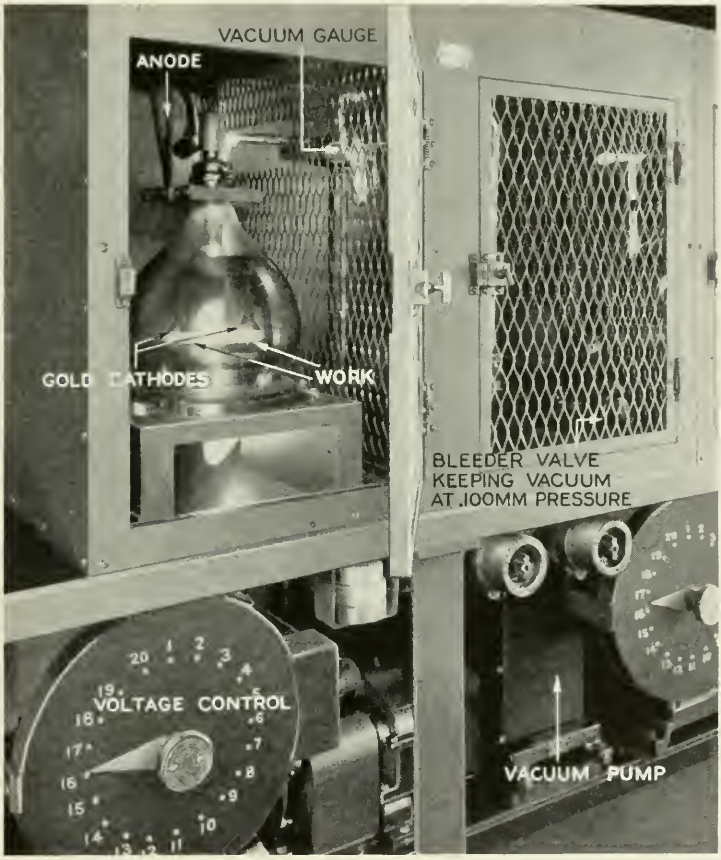


Fig. 3—A complete sputtering unit.

In order to maintain a constant residual gas pressure, the pump is operated continuously and air is allowed to leak in slowly through the bleeder valve which is located near the pump. This practice was found necessary in order to overcome variations in pressure due to the early evolution of gases and the later cleanup usually accompanying electrical discharges in vacuo. The valve is of rugged construction as shown in Fig. 5 and consists of a standard No. 0 taper pin about $2\frac{1}{2}$ inches long, very closely lapped into a bronze bushing. A pressure of .100 ± .005 mm. is readily maintained by this method. After a new charge has been placed in the bell jar, the bleeder valve is temporarily cut off by closing a stop cock so that the required vacuum can be more quickly obtained. By this means, sputtering can be started in about four minutes after the bell jar has been placed in position.

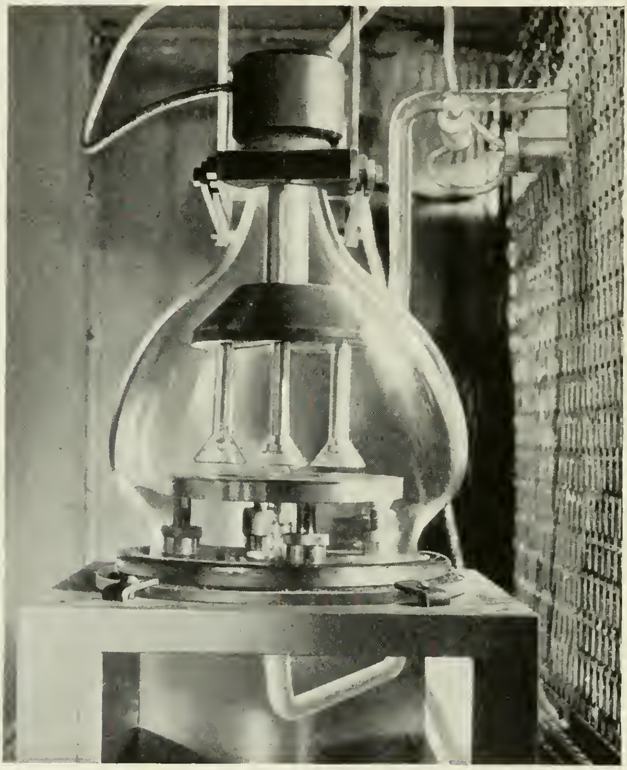


Fig. 4—The sputtering chamber, showing multiple electrodes, diaphragm holder and shield.

The discharge is produced by means of a 1/2 kva transformer which steps up the voltage from 110 to 10,000 volts and is regulated by means of a rheostat in the primary. Safety for the operator is insured by having all high tension leads and terminals enclosed in an expanded metal cage. The door operates a switch in the primary circuit, so that as the door is opened, the circuit is broken.

In order that equal amounts of metal might be evenly deposited on each of the six diaphragms, it was necessary to determine experimentally the relative spacing of the electrodes, and the shape and the size of the shields, stands, and bell jar which would give a uniform current distribution. The best results were obtained by placing the diaphragms 5/8″ from the cathodes, which are arranged symmetrically, by maintaining the pressure at .10 ± .05 mm., by having a 1/2″ hole in the center of the diaphragm holder and shield, and a one inch clearance between their outer rims and the bell jar.

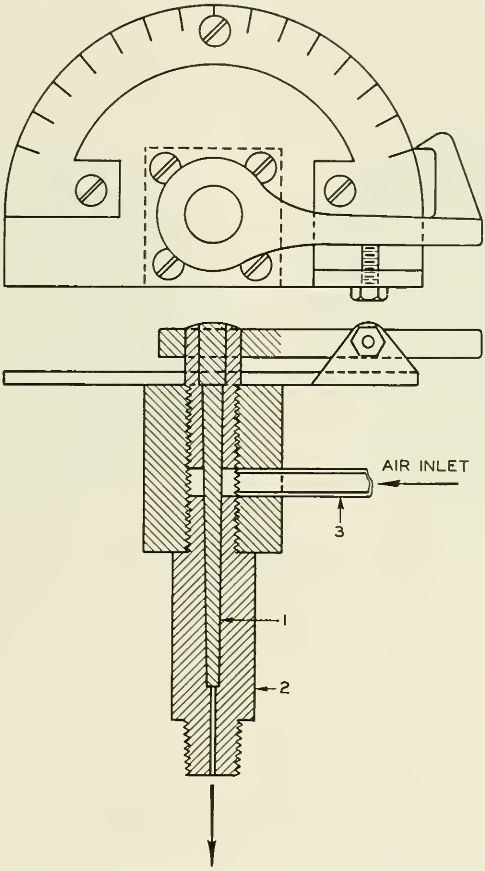


Fig. 5—Detailed construction of the bleeder valve.

In commercial practice a single operator runs three units simultaneously and produces about 90 broadcasting microphone diaphragms a day. These diaphragms require a gold spot on both sides which necessitates breaking the vacuum and turning the parts over. The time not occupied in loading the fixtures and operating the equipment is taken up in preparing diaphragms for sputtering, marking, and packing the finished product. It requires about forty minutes to deposit from three to five mg. of gold per spot. Although this deposit is not more than .001 mm. thick, it is very continuous, adherent, and altogether suitable as a microphonic electrode surface (Fig. 6).

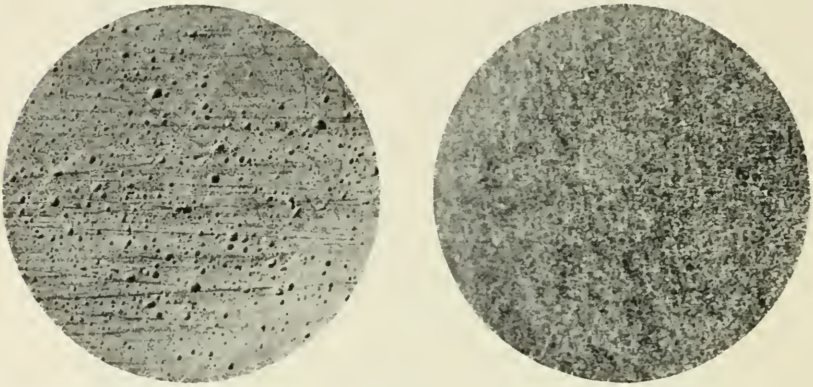


Fig. 6—Typical magnified gold surfaces after three months' service. The spot on the left was produced by electrolytic plating and shows many blisters; the one on the right was produced by sputtering.

This equipment has been in successful operation for a number of years in the production of microphone diaphragms.

Bibliography

W. R. Grove, *Phil. Trans.*, 1852, I.
J. P. Gassiot, *Proc. Roy. Soc.*, **9**, 146, 1858.
Plucker, *Pogg. Ann.*, **103**, 90, **104**, 116, **105**, 70, 1858.
Gassiot, *Rep. of Brit. Assoc.*, **2**, 38, 1861; *Fortschr. der Phys.*, II, 505, 1862.
Wright, *Amer. Journ. of Sc.* (Silliman), **13**, 49; **14**, 169, 1877.
Puluj, *Wien. Ber.*, **81**, II, 864, 1880.
Fr. Wachter, *Wied. Ann.*, **17**, 903, 1882.
E. Wiedemann, *Wied. Ann.*, **20**, 756, 1883.
Hittorf, *Wied. Ann.*, **21**, 126, 1884.
A. Kundt, *Wied. Ann.*, **27**, 59, 1886 und B. Dessau, *Wied. Ann.*, **29**, 353, 1886.
W. Crookes, *Proc. Roy. Soc.*, **50**, 88, 1891.
Moser, *Wied. Ann.*, **42**, 639, 1891.
G. Granquist, *Ofvers. Svensk. Vetensk. Ak. Forh.*, **54**, 575, 1897.
G. Granquist, *Ofvers. Svensk. Vetensk. Ak. Forh.*, **55**, 709, 1898.
Wehnelt. *Wied. Ann.*, **67**, 425, 1899.
A. C. Longdon, *Phys. Rev.*, **11**, 40, and 84, 1900.
J. Patterson, *Cam. Phil. Soc. Proc.*, **11**, 118, 1901.
J. J. Thompson, *Cam. Phil. Soc. Proc.*, **11**, p. 120, 1901.

Goldstein, *Verh. d. D. Phys. Ges.*, 1902, 237.
W. Wien, *Phys. Zeitschr.*, **3**, 440, 1902.
G. C. Schmidt, *Ann. d. Phys.*, **9**, 703, 1902.
J. Tafel, *Ann. d. Phys.*, **11**, 613, 1903.
L. Houllevigue. *Jour. de Phys.*, **4-S**, 2, 1903.
G. C. Schmidt, *Ann. d. Phys.*, **13**, 622, 1904.
J. Tafel, *Ann. d. Phys.*, **14**, 206, 1904.
L. Holborn und L. Austin, *Wissensch. Abhandl. d. Phys.-techn. Reichsanstalt*, Bd. IV, 101, 1904.
S. Valentiner und R. Schmidt, *Ann. d. Phys.*, **18**, 187, 1905.
J. J. Thomson, *Proc. Cambridge Philos. Soc.*, XIII, 212, 1905; *Chem. C.-B.*, I, 891, 1906.
Ch. Maurain, *C. r.* **141**, 1223, 1905.
J. Stark, S. 433, "Electricity in Gasses."
J. J. Thomson, S. 468, "Conduction through Gases."
W. Ramsay, *Proc. Roy. Soc.*, **77**, 148, 1906; *Zeitschr. f. Phys. Ch.*, **55**, 537, 1906.
V. Kohlschutter und R. Muller, *Zeitschr. f. Elektroch.*, **12**, 365, 1906.
V. Kohlschutter, *Zeitschr. f. Elektroch.*, **12**, 869, 1906.
L. Houllevigue, *Jour. de. Phys.*, **6**, 596, 1907.
V. Kohlschutter und Th. Goldschmidt, *Zeitschr. f. Elektroch.*, **14**, 221, 1908.
Th. Goldschmidt, Inaug.-Dissert. Strasburg i.E. 1908.
Fr. Fischer und O. Hahnel, *Zeitschr. f. Elektroch.*, **14**, 366, 1908.
V. Kohlschutter, *Zeitschr. f. Elektroch.*, **14**, 417, 1908.
Fr. Fischer und O. Hahnel, *Zeitschr. f. Elektroch.*, **14**, 433, 1908.
V. Kohlschutter, *Zeitschr. f. Elektroch.*, **14**, 437, 1908.
V. Kohlschutter, *Zeitschr. f. Elektroch.*, **14**, 681, 1908.
B. Walter, *Z. Electrochem.*, **14**, 695, 1908.
F. Fisher, O. Haehnel, *Z. Electrochem.*, **14**, 677, 1908.
Fr. Fischer, *Zeitschr. f. Elektroch.*, **15**, 512, 1908.
J. Stark, *Zeitschr. f. Elektroch.*, **14**, 754, 1908.
J. Stark, *Zeitschr. f. Elektroch.*, **15**, 509, 1909.
J. Stark, Univ. Berlin, *Z. Electrochem.*, **15**, 1909.
Pfund, *Phys. Rev.*, **28**, 324, 1909.
J. Stark, *Z. Electrochem.*, **15**, 509, 1909.
V. Kohlschutter, *Z. Electrochem.*, **15**, 316, 1909.
L. Houllevigue, *Compt. Rendus*, **148**, 1320, 1909 and 1237, 1910.
V. Kohlschutter, *Z. Electrohem.*, 930, 1910.
V. Kohlschutter, *Zeitschr. f. Elektroch.*, **17**, 393, 1911.
H. V. Dechend und W. Hammer, *Zeitschr. f. Elekroch.*, **17**, 235, 1911. (*Ber. d. Naturf. Ges. zu Freiburg i.B.*, **18**, 127); *Vgl. auch Jahrb. d. Rad. und Elektr.*, **8**, 83, 1911.
Kohlschutter, *Jahrbuch. d. Radioakt.*, **9**, 355, 1912.
H. B. Williams, *Phys. Rev.*, **2**, 402, 1913.
O. Steehlman, *Phys. Rev.*, **6**, 330, 1913.
C. W. C. Kaye, *Proc. Phys. Soc. Lon.*, **25**, 198, 1913.
C. Bergholm, *Am. Physik*, **43**, 1–23, 1914.
W. Planck, *Phys. Zeit.*, 563, 1914.
W. F. G. Swann, *Phil. Mag.*, **28**, 467, 1914.
A. M. Tyndall and H. G. Hughes, *Phil. Mag.*, **27**, 415, 1914.
B. Pogany, *Phys. Zeit.*, **15**, 688, 1914.
Nicholson, *Phys. Rev.*, **3**, 1, 1914.
H. B. Williams, *Phys. Rev.*, **4**, 517, 1914.
B. Pogany, *Phys. Zeit.*, **17**, 251, 1916.
W. Reinders and L. Hamburger, *K. Akad. Amst. Proc.*, **19**, 958, 1917.
R. W. King, *Phys., Rev.*, **10**, 292, 1917.
Fabry and Bursson, *Jour. de. Phys.*, June, 1919.
J. A. Becker, L. F. Curtiss, *Phys. Rev.*, **15**, 457, 1920.
J. J. Thompson, "Rays of Positive Electricity," p. 121, 1921.
H. Kahler, *Phys. Rev.*, **18**, 210, 1921.
F. Gross, *Z. Phys.*, **6**, 376, 1921.
Kingdon and Languiuir, *Phys. Rev.*, **20**, 107, 1922.
F. Gross, *Z. Phys.*, **16**, 40–5, 1923.
K. H. Kingdon, *Phys. Rev.*, **21**, 210, 1923.

Kingdon and Langmuir, *Phys. Rev.*, **22**, 148, 1923.
Res. Staff G. E. (London), *Phil. Mag.*, **45**, 98, 1923.
K. Lauch, *Ann. d. Phys.*, **74**, 55, 1924.
S. S. MacKeown, *Phys. Rev.*, **23**, 85, 1924.
L. R. Ingersoll and S. S. de Vinney, *Phys. Rev.*, **26**, 86, 1925.
J. Mazur, *Bull. Acad. Polonaise*, **81**, 1925.
H. Dewhurst, *Phys. Soc.*, **39**, 39, 1926.
A. Guntherschutze, *Zeit. f. Phys.*, **36**, 8, 563, 1926.
L. A. Fourdes, *Ind. Assoc. Cult. Sci. Proc.*, 323, 1926.
A. Guntherschutze, *Zeit. f. Phys.*, **38**, 8, 575, 1926.
E. Blechaschmidt, *Ann. d. Phys.*, **81**, 999, 1926.
A. V. Hippel, *Ann. d. Phys.*, **80**, 7, 672, 1926.
F. Joliot, *Compt. Rend.*, **196**, 1926.
"Cathode Disintegration," Engineering, p. 410, 1927.
F. W. Warburton, *Phys. Rev.*, **30**, 673, 1927.
B. Steinberg, *Phys. Rev.*, **29**, 352, 1927.
A. Andant, *Rev. d'Optique*, **6**, 364, 1927.
A. Andant, *Rev. d'Optique*, **6**, V. 81, 1043, 1927.
T. Baum, *Zeit f. Physik*, **40**, 9, 686, 1927.
A. Ferry, *Jour. de. Phys.* **9**, 39, 1928.
A. Ferry, *Compt. Rend* , **187**, 1928.
H. Nagaoka and T. Fretagami, *Imp. Acad. Tokyo Proc.*, **4**, p. 20, 1928.
L. R. Ingersoll and L. O. Sordahl, *Phys. Rev.*, **32**, 649, 1928.
R. S. Bartlett, *Phyl. Mag.* **5**, 848, 1928.
L. R. Ingersoll and J. D. Hanawalt, *Phys. Rev.*, **34**, 972, 1929.
T. Asada and K. Quasaborth, *Zeit. f. Phys. Chem.*, **143**, 5–6, 1929.
Sophie Dembenska, *Zeit. f. Phys.*, **54**, 1 and 2, 1929.
F. Smith, U. S. P., **1**, 710, 747, 1929.
A. Knapp, *Am. Cer. Soc. Jour.*, **13**, No. 11, 1930.
L. R. Ingersoll, *Nature*, 126, 204, 1930.
L. Van Atta, *Rev. Sci. Inst.*, **1**, 687, 1930.
F. Joliot, *Compt. Rend.*, **190**, 627, 1930.
E. Perucca, *Ann. d. Phys.*, **4**, 2, 252, 1930.
C. H. Cartwright, *Rev. Sci. Inst.*, **1**, 758, 1930.
F. Joliet, *Ann. d. Phys.*, **15**, 418, 1931.
Klemensiewicz and Waswicz, *Zeit. f. Phys.*, **71**, 11 and 12, 817, 1931.

A Voice and Ear for Telephone Measurements

By A. H. INGLIS, C. H. G. GRAY and R. T. JENKINS

An artificial voice and ear have been developed which are sufficiently close simulations of the real voice and ear in their principal physical attributes to justify their use in both shop and laboratory tests of telephone transmitters and receivers.

The artificial voice and ear have certain advantages in that they can be exactly specified and reproduced and can be used in determining physical characteristics of instruments which are difficult or impossible to obtain with real voices and ears.

THE performance of telephone transmitters and receivers is dependent not only upon their physical characteristics but also upon the reactions of the users to these instruments. Observations of the results obtained by subscribers under known conditions of practical telephone use take both these factors into consideration. Tests of this kind, however, are not suited to the needs of laboratory development and study of instruments where data regarding their physical characteristics are required. In such instances data, to be of greatest value, should be taken under conditions which include those factors of actual service having an important bearing on the performance of the instruments.

Among the most important of such factors are the voice and ear. For making many of the laboratory tests, therefore, it has been necessary to employ actual human voices and ears in order to insure that all of their physical characteristics have been included in the test. This procedure, however, due to uncontrollable variations in individuals, requires a large expenditure of time and effort to insure the precision desired. Furthermore, certain instrument tests, such as response-frequency characteristics, are either impossible or difficult to make with the real voice or ear.

These disadvantages, inherent in the use of the human voice and ear, have been recognized for a long time. Numerous attempts have been made to employ voice and ear substitutes for instrument testing, and in cases where uniformity of instruments rather than their design has been of primary importance, as in shop acceptance tests, such substitutes have been of great value. That their use in engineering design problems has not been more extensive has been due to the inability to make them meet certain fundamental requirements.

It has been the aim, therefore, in the design of the artificial voice and ear to be described, to overcome previous objections to the use of such substitutes so that they may be used with confidence in general testing and physical measurement of telephone transmitters and receivers. The requirements which an artificial voice and ear must meet to insure proper performance are outlined below.

Requirements for an Artificial Voice

An ideal artificial voice must be able to reproduce human speech without introducing any change in frequency, amplitude or directivity over the entire range of intensities possible in speech. Furthermore, it should react to changes in acoustic load in the same manner as does the human voice under ordinary conditions. Such requirements, of course, must be reduced to a more specific and practical form to enable the construction of a physical piece of apparatus.

It is useful to consider the artificial voice as consisting of two parts, the mouth and the source of power. Practical considerations point immediately to the use of some form of electro-acoustic transducer for the mouth, and any of several sources of electrical power. In the production of speech there is required as a source of power either a high quality transmitter or a phonograph record and reproducing system. For purposes of physical measurement and analysis, a source of single-frequency power, such as an oscillator, is needed.

If it be assumed that the frequency composition of the actual human voice is automatically included with either of the sources of speech mentioned, and that proper frequency weighting will be introduced in the single-frequency source when desired by means of suitable electrical networks, the practical requirements for the artificial mouth may then be stated specifically as follows:

(1) It should introduce no amplitude distortion within the range of speech frequencies.
(2) It should be capable of delivering an acoustic output without nonlinear distortion over the range of intensities possible for the human speaking voice.
(3) The distribution of the sound field about the mouth at every frequency and distance should be the same as that of the human mouth.
(4) The introduction of objects such as transmitters in the sound field should react on the output of the mouth and distort the field in the same way as they do when introduced in the field of the human mouth.
(5) It should be completely specifiable and reproducible as well as constant in performance.

Requirements for an Artificial Ear

In considering a substitute for the human ear, the general requirement is that it shall respond, as does the human ear, to sounds of various frequencies, amplitudes and lengths of duration. Furthermore, its reaction on the source of sound, whether that sound be produced by a receiver held against it or by a source at a distance from it, shall be the same as that of the human ear.

As in the case of the artificial voice, the artificial ear may conveniently be considered as consisting of two parts, the ear coupler and the measuring equipment. Based on the above general considerations, the ear coupler should meet the following more specific requirements:

(1) It should have the same impedance at every audible frequency as a real ear, either in the open air or with the receiver held to it.

(2) The pressures developed in the ear coupler should be the same at every audible frequency as the pressure developed in a real ear.

For the measuring equipment, the requirements are as follows:

(1) For steady state conditions it should be capable of giving an indication at every audible frequency proportional to the pressure in the coupler over a range of pressures as great as that experienced by a real ear.

(2) It should respond to sound of short duration as does a real ear.

(3) It should be possible to change the response to various frequencies in any manner required by means of suitable electrical networks.

(4) It should respond to complex sounds as does the real ear.

The artificial ear should be completely specifiable, reproducible, and constant in performance.

At the present time it is not possible to meet rigidly all of these requirements, either for an artificial voice or ear. Development work has progressed to a point where it can be stated that the requirements have been met sufficiently well to enable the production of both an artificial voice and ear which are close simulations of the real voice and ear and which will be satisfactory as substitutes in almost all of the laboratory or shop tests for which a real voice and ear have been largely used.

Description of the Artificial Voice

The schematic arrangement of the artificial voice referred to is shown in Fig. 1. As indicated, electrical energy may be supplied from any one of several sources through the electrical network and amplifier to the electro-acoustic transducer. For single-frequency measurements a heterodyne oscillator provides a convenient means for obtaining the desired testing currents. As a source of speech, advantage can be taken of recent developments in phonograph technique which make it possible to record and reproduce speech practically without distortion.[1] This affords a very satisfactory source of speech for the artificial voice. In certain special instances where it is desired to make measurements with human voices under closely controlled conditions, a high quality transmitter is used as the input element of the artificial voice.

To insure proper operation under both steady-state and transient conditions the electrical network employed in the artificial voice is of the constant-resistance type.[2] In this network is included compensation for deviation from uniformity of response of the electro-acoustic transducer. Additional compensation may be provided in case the input to the artifical voice is from a source of constant output with frequency and it is desired to weight the single frequency output of the artificial voice in accordance with the distribution in speech [3] of pressure with frequency.

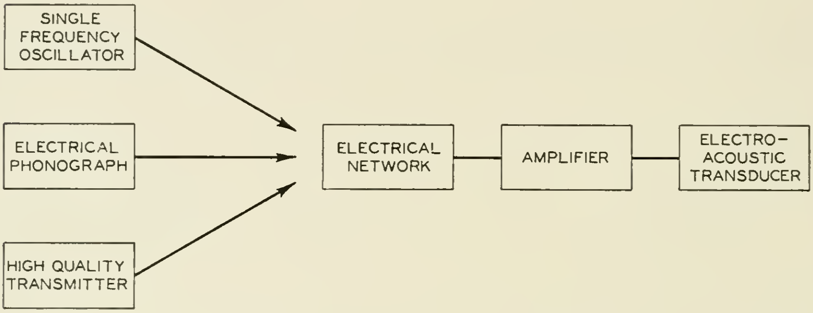


Fig. 1—Schematic arrangement of artificial voice.

The amplifiers employed have high gain and high output capacity. They provide, without introducing distortion, the maximum amount of electrical energy that may be required to obtain the desired acoustical output from the artificial mouth.

[1] "Vertical Sound Records—Recent Fundamental Advances in Mechanical Records on Wax." Presented by H. A. Frederick at Swampscott, Mass., Oct., 1931, before the Soc. of Motion Picture Engineers.

[2] "Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks," Otto J. Zobel, *Bell System Technical Journal*, July, 1928.

[3] "Some Physical Characteristics of Speech and Music," Harvey Fletcher, *Bell System Technical Journal*, July, 1931.

The unit employed as an artificial mouth is shown in Fig. 2 mounted on a stand in such a manner that it may be rotated through an angle of approximately 180° as a matter of convenience in certain types of testing. It is a modification of a loud speaking receiver of large power capacity.[4] As one of the requirements in the design of a substitute for the human voice is that the distribution of the sound field of the artificial voice must be similar to that of the human voice, it is necessary to make the opening or sound radiating surface of the artificial mouth comparable in size with that of the human mouth. To meet this need the horn coupling ordinarily associated with this

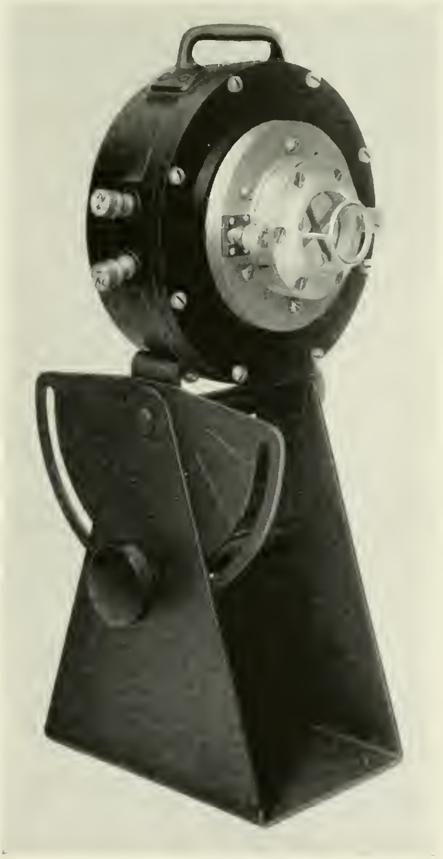


Fig. 2—Artificial mouth.

---

[4] "A High Efficiency Receiver of Large Power Capacity for Horn-Type Loud Speakers," E. C. Wente and A. L. Thuras, *Bell System Technical Journal,* January, 1928.

type of receiver is removed, the throat insert of the receiver modified and a simple structure mounted in place of the horn coupler having an opening effectively that of the human mouth. In this structure is mounted an acoustic-resistance element. This has a mechanical resistance [5] of approximately 41.0 mechanical ohms per cm.^{2} and below 5000 c.p.s. a reactance of less than 10 per cent. There is mounted on the structure replacing the horn coupler, a guard ring which serves as a reference plane for measurements of distance between the artificial mouth and instruments under test. The location of this reference plane has been empirically determined so as to correspond to the plane of the lips of a human mouth.

At low frequencies the radius of the opening of the mouth is small compared to the wave-length. Hence in effect a point source of sound is approached. Under these conditions the radiation resistance is small. As the frequency increases the radiation resistance increases until the wave-length has decreased to a value approximately three times the radius of the opening. At this frequency and above, the radiation resistance is approximately constant at about 41.0 ohms per cm.^{2} The output impedance of the artificial mouth is high with respect to the radiation impedance at low frequencies. Inasmuch as the impedances are not matched except at the higher frequencies the output power should be approximately proportional to the radiation resistance. However, this relationship is modified by the resonances of the instrument. The acoustic resistance reduces these resonances and also serves to reduce the reaction on the artificial mouth which might arise by placing an instrument close to and directly in front of it.

The response-frequency characteristic of the mouth measured at the guard ring is shown in Fig. 3. In order that this response may be uniform over the important speech frequency range of 100 to 7500 c.p.s. the characteristic shown in Fig. 3 is equalized with the network indicated in Fig. 1. The resulting response of the artificial voice is shown in Fig. 4.

Another important requirement is that the artificial mouth shall be capable of delivering without non-linear distortion sound outputs corresponding to what may be termed loud talking for human beings. Because it is desirable to operate at frequencies as low as 100 c.p.s. it is necessary to supply a comparatively large amount of electrical energy to the artificial mouth. To accomplish this an amplifier has been employed as indicated in Fig. 1 which makes it possible to obtain sound pressures of 16 bars at the plane of the guard ring at all frequencies between 100 and 7500 c.p.s. without appreciable harmonics. Speech may be transmitted about 15 db above the average intensity employed in commercial telephone conversations without noticeable distortion. Very few conversations are carried on at an intensity greater than this.

---

[5] "Methods of High Quality Recording and Reproducing of Music and Speech Based on Telephone Research," J. P. Maxfield and H. C. Harrison, *Bell System Technical Journal*, July, 1926, p. 506.

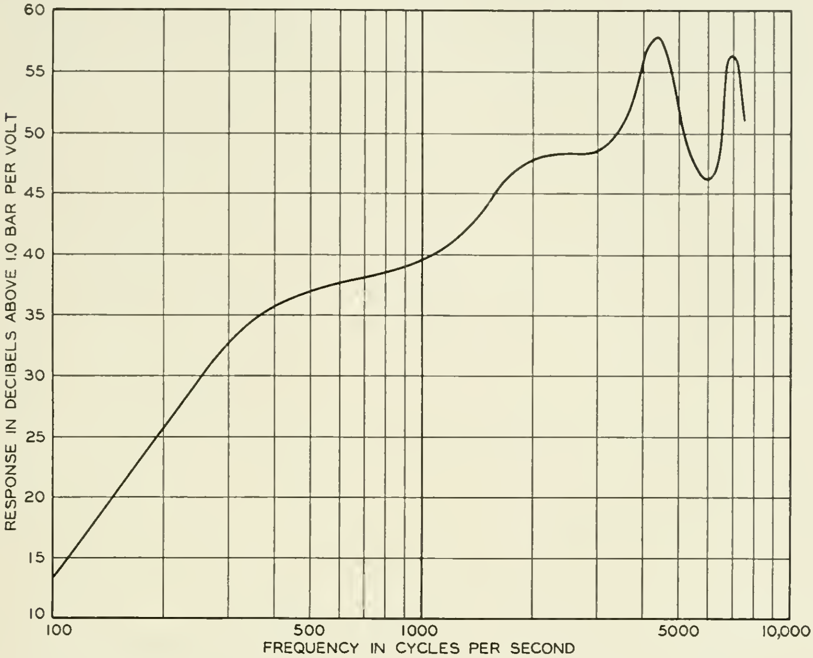


Fig. 3—Response-frequency characteristic of electro-acoustic transducer without equalizer.

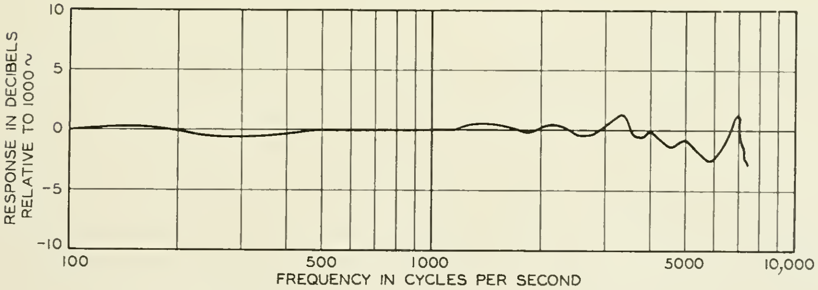


Fig. 4—Response-frequency characteristic of artificial mouth.

Measurements have been made of the magnitude of the harmonics present in the output of the artificial voice under conditions of high sound pressure at 100 c.p.s. where maximum amplitudes of vibration are encountered. Only odd harmonics are present, the greatest in magnitude, the third, being about 15 db below the fundamental at the highest pressure used. At other frequencies or lower intensities, the harmonics are of even less importance. At the average intensity employed in telephone conversations no appreciable harmonics whatever are present. The results of these measurements are shown in Fig. 5.

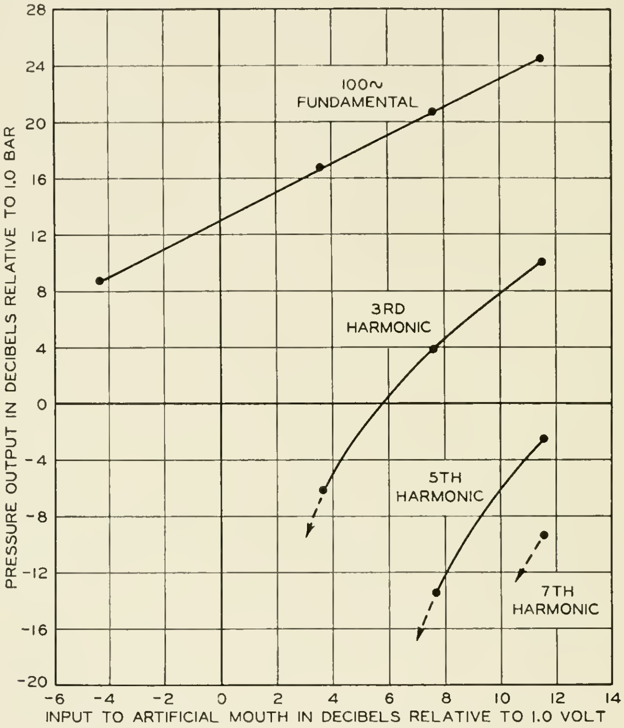


Fig. 5—Harmonics in output of artificial mouth.

It has been mentioned that the sound field distribution of the human and artificial voices should be alike. One method of determining agreement in this respect has been to measure the output of several different types of transmitters when the instruments were placed at various distances from both the artificial mouth and the human mouth. Eight persons, four men and four women, were employed in the actual voice tests, to call the testing phrases, "Joe took father's shoe bench out" and "She was waiting at my lawn." These phrases include all of the important speech sounds and are brief and easy to say. The speech output of these same individuals calling the above phrases was recorded by the process mentioned.[1] After adjusting the output of the artificial voice for the close talking position with a condenser transmitter,[6] tests were made on the various instruments interspersing both instruments and voices to reduce testing errors and minimize changes in instrument characteristics between tests. The results of individual tests on a given type of instrument with both the human and artificial voices varied over approximately the same range. The average results for all tests for each type of instrument for each distance and for both the human and artificial voices are shown in Fig. 6. These data show reasonably close agreement between the artificial voice and human voice, both as regards the general level and slope of the distance loss characteristics. In the practical application of the artificial mouth to telephone instrument studies the minor discrepancies shown may be minimized by the use of correction factors. A further investigation is being made of the sound field distribution of the artificial mouth in comparison with that of human mouths.

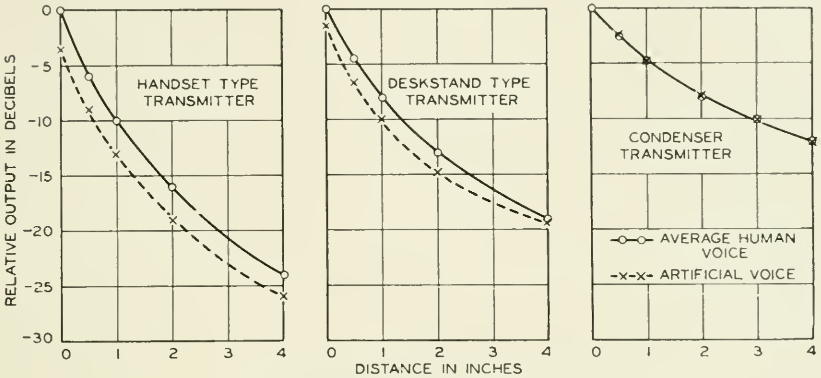


Fig. 6—Distance loss characteristics.

Tests were made to determine the effect on telephone instruments of the size and shape of the artificial mouth. Response-frequency measurements were made, a deskstand and a handset transmitter being used, each modified to include in the plane of the diaphragm a small condenser transmitter. The latter was used in order that any variation of carbon instruments might be eliminated. Sound pressure was obtained from each of four types of artificial mouth. Three of these employed a long pipe with an inside diameter of about 0.7 inch.

[1] Loc. cit.

[6] "Electrostatic Transmitter," E. C. Wente, *Phys. Rev.*, May, 1922.

In one case no baffle or reflecting surface of any kind was used at the pipe opening. In the other cases it was terminated (1) in a replica of a human head molded of soft rubber (2) in an 8-inch square baffle. The fourth type used was the artificial mouth with acoustic resistance shown in Fig. 2. The variation in pressure measured for these various conditions is shown on Fig. 7. It will be noted that except for the extreme condition of the pipe without termination, good agreement is obtained for the various response-frequency characteristics, indicating that the shape of the artificial mouth is not critical provided the opening is of the proper order of magnitude and is effectively in a baffle commensurate with the size of the human head. The measurements described above were made with the plane of the mouthpiece of the instruments in a representative position with respect to the plane

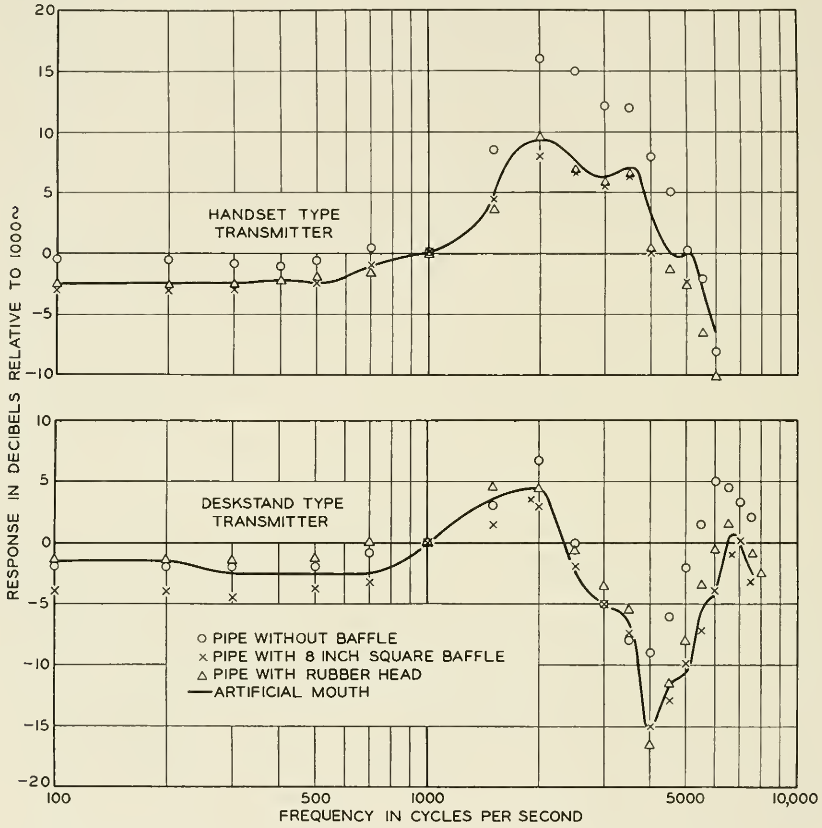


Fig. 7—Effect on station transmitters of variation in size and shape of artificial mouth.

of the mouth opening. To determine the effect of increasing the distance between the transmitter and the mouth, measurements were made with the deskstand type of instrument with results as shown in Fig. 8.

Measurements were made of the distortion of the sound field from each of the four types of mouth mentioned which was caused by the introduction of different types of transmitter in the field. The effects were measured immediately to one side of the mouthpiece. It will be seen from Fig. 9 that for frequencies below 2000 cycles there is good agreement for all types of mouth and both types of instrument. Above 2000 c.p.s. variations appear, but in general the largest differences occur as before for the extreme form of artificial mouth, namely, the pipe without termination.

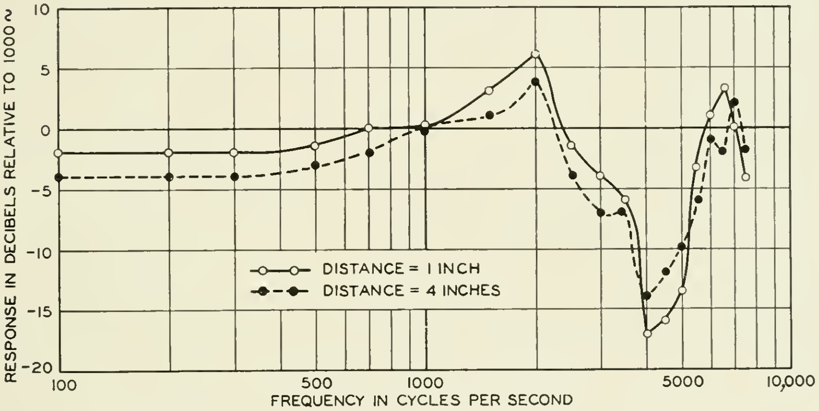


Fig. 8—Effect on station transmitters as distance from mouth is varied.

Summarizing the results obtained with the artificial mouth described, although the general requirements have not been completely satisfied, a rather close approximation has been realized. The response-frequency characteristic does not vary more than ± 1.0 db from 250 to 3000 c.p.s. nor more than ± 2.0 db from 100 to 7500 c.p.s. The mouth is capable of delivering an output of an intensity equivalent to loud talking without appreciable non-linear distortion. The distribution of the sound field is similar to that of the real mouth. As judged by tests on several different forms of mouth of quite different sizes and shapes, the indications are that the introduction of objects in the sound field of the artificial mouth chosen distorts that field in about the same manner as occurs with the real mouth. Speech reproduced by the artificial voice sounds natural. Comparative articulation tests of direct speech from an individual and its reproduction by the artificial voice agree within a few per cent. The design can be definitely specified and reproduced with accuracy and it is rugged in structure and constant in performance.

General experience in the use of this artificial voice over about a year's time has indicated that it may be used satisfactorily in forms of transmitter testing which have heretofore required the human voice.

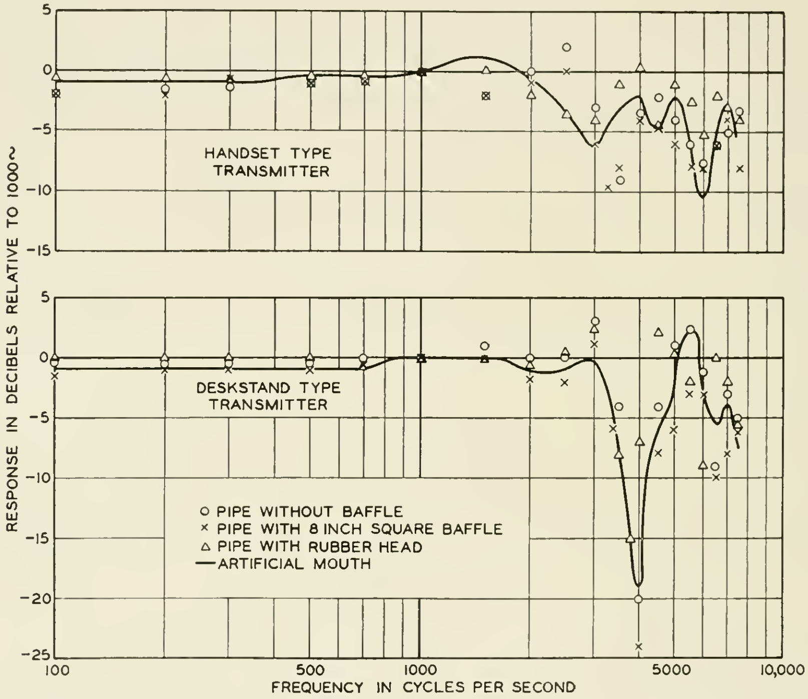


Fig. 9—Effect of introducing station transmitters in sound field of artificial mouth.

The Artificial Ear

The human ear is limited in its utility generally to relative or comparative measurements, and is not well adapted for measurement of absolute pressure or velocity. At best, measurement with real ears is a laborious procedure. In general, it has been found necessary to take a large number of observations with many ears, since a measurement with one ear is unreliable. One of the principal reasons for this unreliability is the difficulty not only of securing a particular coupling of a receiver held to the ear, but of duplicating this coupling in subsequent measurements.

The output of a receiver coupled to the ear is governed by several factors, among which are: enclosed volume, leakage around the cap, constriction of the ear canal and the yielding of walls and tympanum. Resonances in the enclosed volume are also generally present, associated with the dimensions of the chamber. The effects of the acoustic load of an ear coupled to a receiver are, in general, increased damping, higher resonant frequency and a dropping off of the receiver response at low frequencies. The receiver response also shows peaks due to resonances in the enclosed chamber.

Largely as a matter of convenience, couplers of simple construction imposing a stiffness load have been frequently used for receiver calibrations. It has been recognized that calibrations made in this way do not agree with the performance realized when the receiver is held to the ear and this has led to the development of an artificial ear which will permit the measurement of receivers of any type under conditions closely simulating those under which they are used.

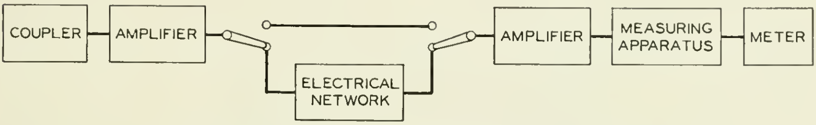


Fig. 10—Schematic arrangement of artificial ear.

The schematic arrangement of this artificial ear is shown in Fig. 10. It comprises: a special coupling device designed to present the same acoustic load to a receiver as does a typical human ear, a small condenser transmitter serving as the measuring element in the coupler, and means for amplifying and measuring the voltages generated by the condenser transmitter. For certain purposes arrangements are provided for introducing an electrical network of the constant-resistance type [2] having a response-frequency characteristic corresponding to that of the equal loudness curve [7] of human hearing at the desired sensation level.

The coupler is shown in Figs. 11 and 12. The cap of the receiver under test rests upon a molded soft rubber insert which has the internal contour of the auricle. Soft rubber was selected because of its yielding qualities which are of importance in two respects: it permits the receiver under test to be sealed to the coupler without the aid of such substances as petroleum jelly, and it can be readily molded to have the desired shape. In this rubber insert is an acoustic leak consisting

---

[2] Loc. cit.

[7] "A Direct Comparison of the Loudness of Pure Tones," B. A. Kingsbury, *Phys. Rev.*, April, 1927.

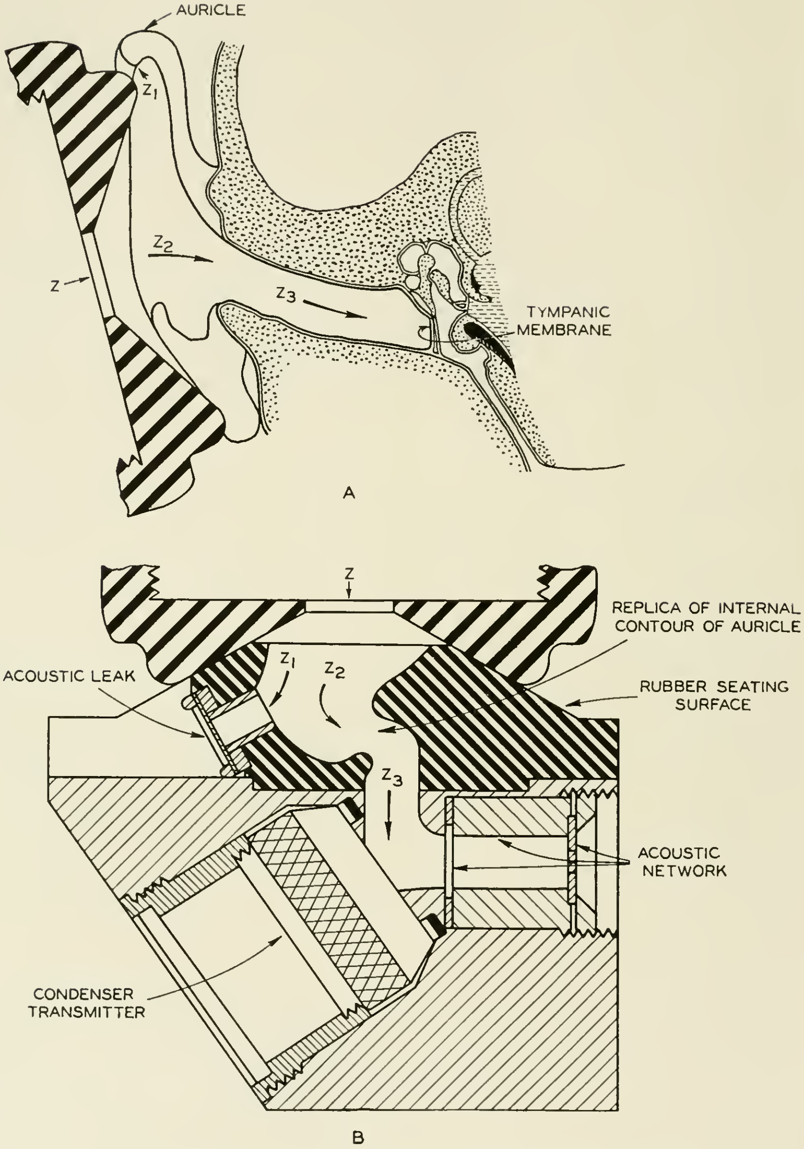


Fig. 11—Sectional views of human ear and artificial ear coupler.

of a tube terminated by an acoustic resistance. The purpose of this leak, which has both mass and resistance, is to simulate the leak between a receiver cap and a typical human ear. The molded rubber insert is attached to a rigid structure in which a chamber has been provided to simulate the size and, to some extent, the shape of the auditory canal of a human ear. The diaphragm of the small condenser transmitter has been placed in the wall of this chamber to measure the pressures developed. The chamber is terminated by an acoustic network having mass, stiffness and resistance components of

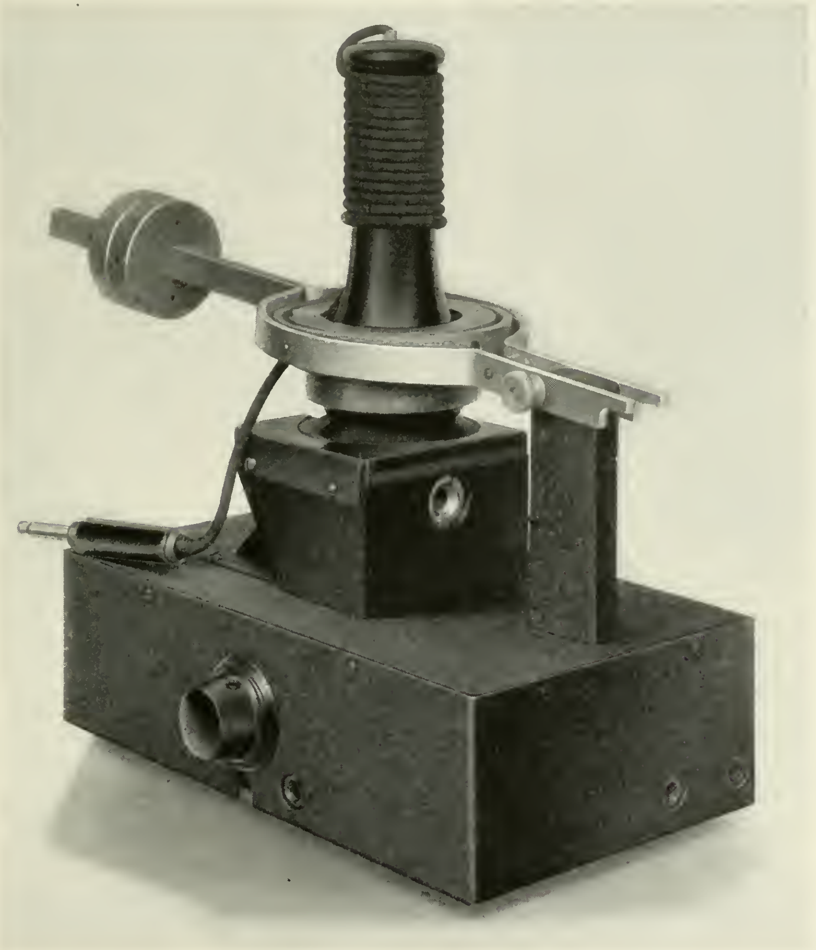


Fig. 12—Artificial ear coupler.

the same order of magnitude as those observed looking into the auditory canal of a human ear. In considering Fig. 11, it should be noted that, due to the irregular contour of the auricle, one sectional drawing cannot adequately portray the shape of the opening in the molded soft rubber insert.

The box shown in Fig. 12, in which the artificial ear is mounted, houses a two-stage amplifier for use with the small condenser transmitter. The mechanical structure around the receiver is used for centering it on the coupler and to hold it in position at a definite pressure, the force, of course, always exceeding the weight of the receiver.

The magnitude of the impedance (Z) indicated in Fig. 11 was measured looking through the aperture of a conventional type of receiver cap held in a normal manner to the ear by each of 14 men. As might be expected, wide variations were observed between individual ears, particularly for the lower frequencies. Considerable variation was also observed on repeated tests on an individual. Table I shows the magnitude and range for both the resistance and reactance components of the acoustic impedance observed in the measurements made at several frequencies. Supplementing these data is Fig. 13 which shows the acoustic resistance and reactance for typical human ears with and without a leak between the receiver cap and the ear, together with similar data on the artificial ear. It will be noted that the various impedance curves for comparable conditions of the human and artificial ears are quite similar in shape. There is, however, some discrepancy in the magnitudes of the impedance for comparable

TABLE I

Acoustic Impedance of Ears Looking through Aperture of Receiver Cap

f	Observed Range of Measurements for 14 Male Human Ears		Artificial Ear			
	Resistance	Reactance	Resistance		Reactance	
			(1)	(2)	(1)	(2)
100....	1 to 70	−300 to +24	21	16	−295	+15
300....	1 to 80	−195 to +60	14	56	−112	+59
400....	1 to 200	−115 to +92	11	149	− 72	+ 5
800....	1 to 107	−111 to +10	15	15	− 45	−45
1200....	3 to 18	− 36 to +30	11	11	− 24	−24
2300....	1 to 21	− 21 to −5	5	5	− 15	−15

(1) With no leak.
(2) With typical leak between receiver cap and ear.

conditions. In view of the large range in the impedance values obtained for different individuals and at different times the discrepancies indicated are of doubtful importance. In this connection it should be noted that a wide range in impedance values could be obtained readily on the artificial ear merely by changing the constants of the acoustic networks. In view of the fact that a fixed condition of the artificial ear is desirable for many practical purposes the chief requirement as regards the impedance characteristics under discussion is that they closely approximate the impedance characteristics of a typical ear.

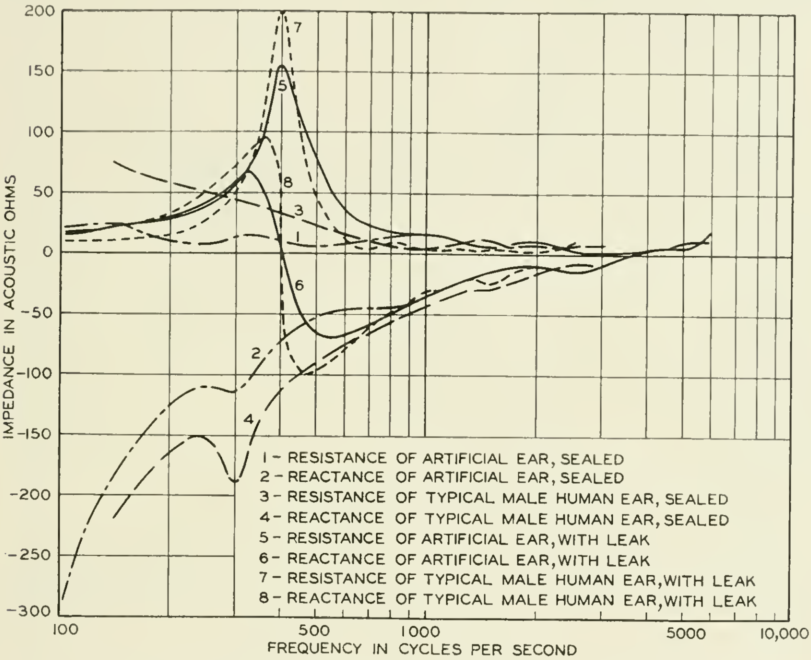


Fig. 13—Acoustic impedance of ears as viewed through aperture of receiver cap.

Considerably more data on human ears than are now available appear to be required before an artificial ear with more typical impedance characteristics than those shown can be designed.

In Fig. 14 additional acoustic impedance data on a typical male human ear and on the artificial ear are shown, the measurements in this case being made looking into the auditory canal. Table II presents data supplementing Fig. 14. As in the previous measurements, wide variations were encountered between different ears at any given frequency. The agreement between the artificial ear

TABLE II

Acoustic Impedance of Ears Looking into Auditory Canal

f	Observed Range of Measurements for 7 Male Human Ears		Artificial Ear	
	Resistance	Reactance	Resistance	Reactance
200....	50 to 250	−1134 to −633	60	−610
400....	20 to 250	− 760 to −254	80	−385
800....	35 to 136	− 300 to −130	15	−145
1200....	60 to 91	− 180 to −20	20	− 80
2000....	8 to 70	− 155 to −50	20	− 27
3000....	15 to 60	− 140 to +50	22	+ 15

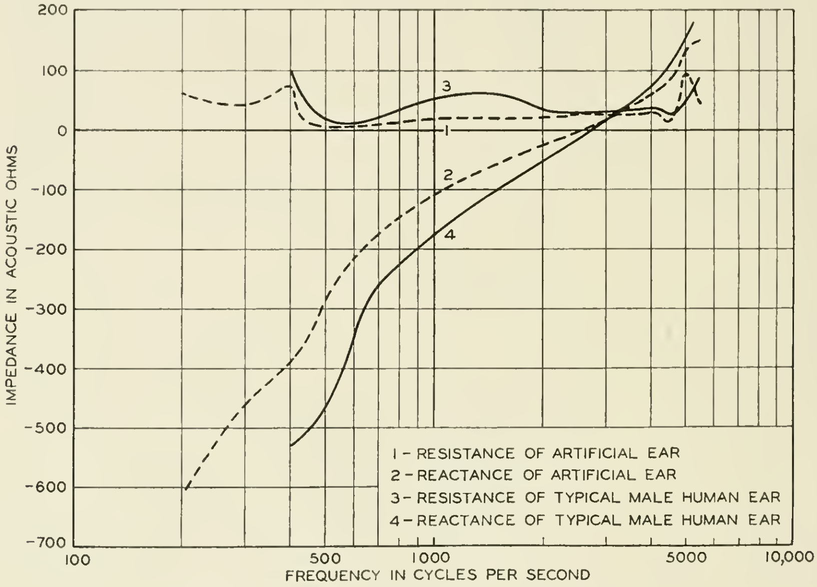


Fig. 14—Impedance characteristic of auditory canal.

results and the human ear measurements is as close as seems warranted by the available data.

The acoustic impedance data presented are important from the standpoint of insuring that the receiver under test on the artificial ear operates under nearly the same load conditions as it does on a human ear. It is also important that the response-frequency characteristic of a receiver obtained on the artificial ear compare well with that obtained on the human ear. In this connection two widely different types of receiver have been studied; one, a moving coil receiver such as is used in the Master Reference System for Telephone Transmission,[8] and the other, a deskstand type receiver. Measurements of the response characteristics of these two types of receiver have been made on male human ears as described in the Appendix. The results obtained, together with similar measurements on the same receivers on the artificial ear, are shown in Figs. 15 and 16. Over most of the frequency range rather good agreement is found between the human ear and artificial ear measurements, the discrepancies in most cases being relatively unimportant.

Electrical impedance measurements were made on two commercial types of receiver whose impedances at resonance are relatively sensitive to changes in acoustic load. These measurements were made with the receiver in free air and when held to the human ear and to the artificial ear. The results are shown in Table III. The values of

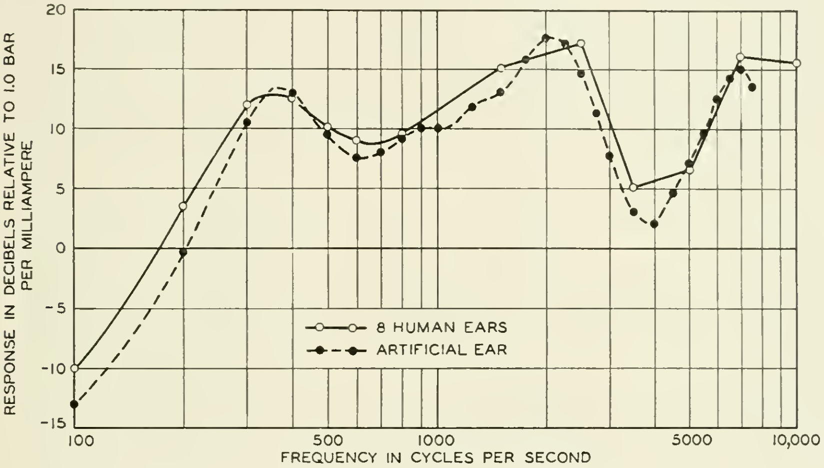


Fig. 15—Response-frequency characteristic of moving coil receiver on ears.

TABLE III

Electrical Impedance Data for Different Types of Receivers with Various Loads

Load Condition	Type of Receiver	Natural Frequency (f_0)	Damping Constant (Δ)
Air	Deskstand	854	122
Artificial Ear	"	1009	208
Observed Range on Male Human Ears	"	777–1062	141–258
Air	Handset	808	122
Artificial Ear	"	967	355
Observed Range on Male Human Ears	"	818–1048	226–524

---

[8] "Master Reference System for Telephone Transmission," W. H. Martin and C. H. G. Gray, *Bell System Technical Journal*, July, 1929.

natural frequency (f_0) and damping (Δ) as obtained from motional impedance circles for the measurements on human ears vary over a wide range. The results obtained on the artificial ear closely approximate those obtained on what may be considered a typical normal male ear.

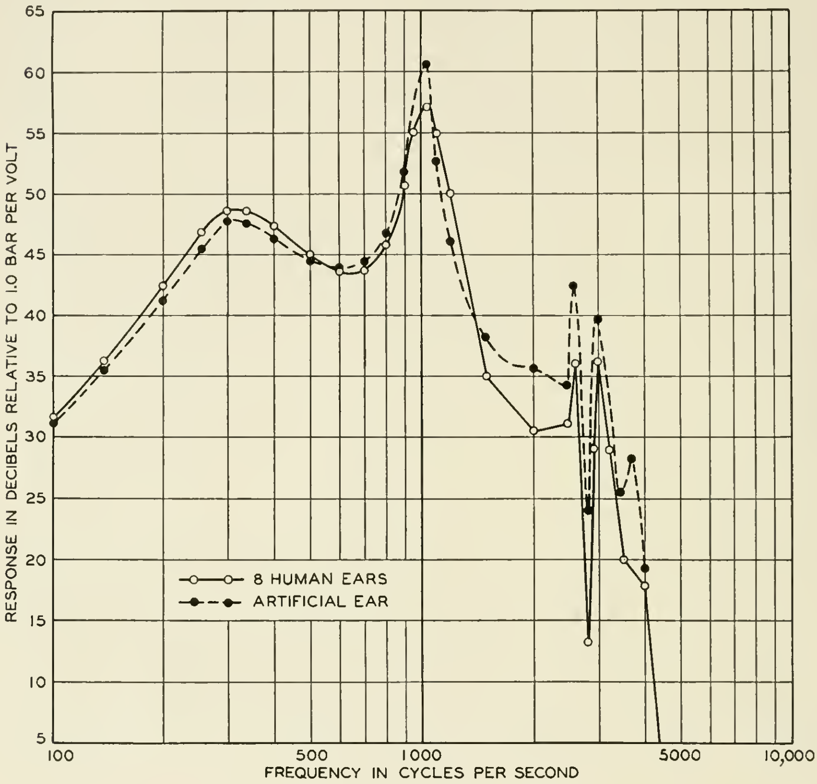


Fig. 16—Response-frequency characteristic of deskstand receiver on ears.

The artificial ear coupled through a suitable amplifier is terminated in measuring apparatus similar to that employed in the sound meter.[9] The indicating instrument used responds to impulses of short duration (under .2 second) in a manner approximating the response of the actual ear to sounds of similar duration. The rectifier, of the copper-oxide type, obeys over its useful range, essentially a square law. By the addition of a suitable loudness weighting network similar to that used in the sound meter, the artificial ear may be used to obtain a measure of the relative loudness produced by different receivers. Other weighting networks may be used to enable a direct meter reading of other desired characteristics.

---

[9] "Indicating Meter for Measurement and Analysis of Noise," T. G. Castner, E. Dietze, G. T. Stanton, and R. S. Tucker. Presented at the Northeastern District Meeting of the A. I. E. E., Rochester, N. Y., April 29–May 2, 1931.

While any single design of artificial ear can, of course, simulate only a single ear condition, there appears to be no reason why a structure of the type described cannot be made to simulate any ear by proper changes in the dimensions of the passage and values of the acoustic impedances without sacrifice of stable performance or the ability to specify and reproduce it. As far as data on the characteristics of human ears are at present available, the particular design described gives a good approximation, both in its effect on the receiver and in its own frequency characteristics, of a typical male ear. There is indication that this artificial ear, in addition, may offer an equally satisfactory substitute for the real ear in the measurement of open sound fields. A further study of this possibility is being made.

General Applications of the Artificial Voice and Ear

Equipment of the kind described offers many advantages, both in aboratory investigation and in shop testing of instruments, as compared with the methods that have been used heretofore. The artificial voice and ear described have, of course, the advantages of exact specification and control of the testing conditions, and of rapidity in obtaining data which have been the principal arguments for the use of previous voice and ear substitutes. In addition, instruments tested by these new means are under nearly normal operating conditions.

In laboratory investigations and development of instruments the artificial voice, makes it possible to carry out tests, such as response-frequency measurements with the same instrumentalities as are used for speech tests. This is, of course, impossible with the real voice. When used in conjunction with a high quality transmitter it permits either the variation of talking intensity without change in quality, or the maintenance of a constant output intensity from the artificial mouth, even though the actual speaker's voice may be varied over a wide range. Applications of this kind in exercising over a caller's voice, control of which he himself is incapable, are invaluable in many laboratory investigations.

As applied to shop inspection practices, the desirability is obvious of having a single testing means for all transmitters and receivers regardless of type, which is identical in principle with the means of testing used in the laboratory. The results of measurements made on instruments, whether in laboratory, factory or repair shop, can be directly compared and the results used to great advantage in the engineering and maintenance of the telephone plant.

A further valuable application of the artificial voice and ear is as an adjunct to a reference telephone system. At the present time uncontrollable differences in technique and testing personnel at various points where it may be desirable to employ such reference systems involve discrepancies difficult to eliminate or explain. The use of the artificial voice and ear with suitable phonograph records makes it possible to have identical testing means at any point desired.

Further experience with the artificial voice and ear will undoubtedly open up new possibilities and applications for these instrumentalities and enable more accurate investigation of instruments with less expenditure in effort and time.

APPENDIX

Response-Frequency Characteristics of Receivers on the Human Ear

In Figs. 15 and 16 are shown response-frequency characteristics of receivers on the human ear. One method by which such characteristics may be determined is shown schematically in Figs. 17-*A* and 17-*B*. With the auricle projecting through an aperture (provided for purposes of definite location of the ear) toward the sound source as shown, the pressure at each frequency produced at the tympanum or as near to it as possible is measured by means of the calibrated transmitter and search tube.

With the pressure measured and the search tube removed from the ear, the observer listens to the sound from the source, as shown in Fig. 17-*Ba*. Then he listens to the receiver to be calibrated as shown in Fig. 17-*Bb*. The electrical input to the receiver is adjusted until the observer judges the sensation to be equal to that from the source. This then gives the pressure produced by the receiver for a given frequency and for a given input to the receiver.

The Search Transmitter and Its Calibration

The purpose of the search transmitter is to furnish an instrument of small external dimensions to meet the following requirements:

1. It must admit of calibration at single frequencies in terms of electrical output per unit pressure at the mouth of the tube.
2. It must be small enough to admit of insertion into the ear canal without material distortion of the sound field in the latter.

The tube used in the measurements is associated with a condenser transmitter.[6] One end of this tube, having a relatively large opening is acoustically coupled to the transmitter diaphragm. The calibration of the transmitter is made with the aid of a Rayleigh Disc, in a highly damped sound-proof box, as shown in Fig. 18. The sound source is a loud speaking receiver of large power capacity,[4] terminated with an adjustable tube. With the mouth of the tube as origin, an approximately spherical progressive sound wave is produced. From the disc deflection, the particle velocity and hence the alternating sound pressure is determined in the space occupied by the disc. The search transmitter (the bulk of it wrapped in felt) is placed with its mouth as near the disc as possible, and its voltage output is measured.

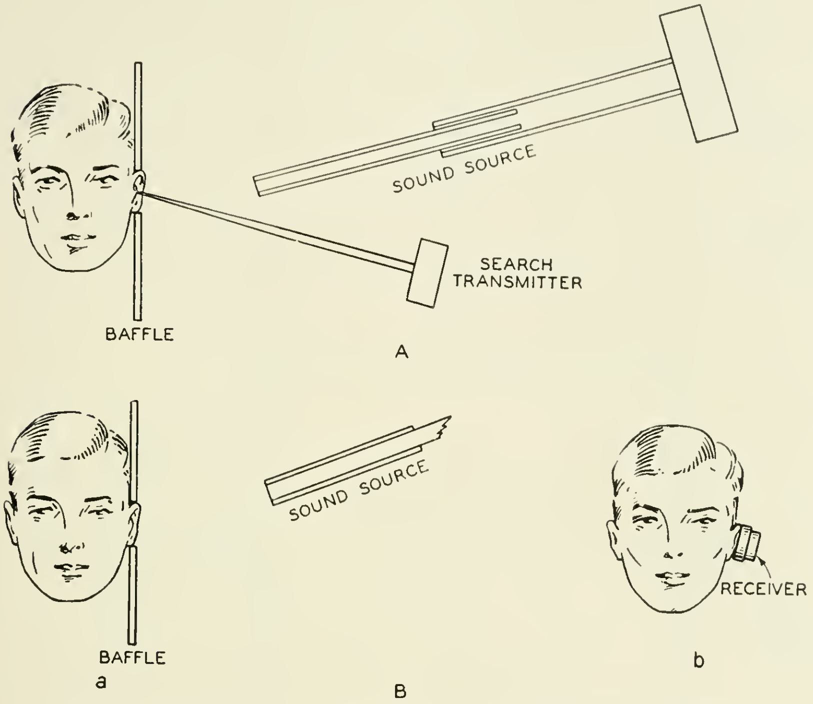


Fig. 17—Diagrammatic arrangement employed in receiver calibrations on human ears.

The search transmitter calibration is made under conditions conforming to the following essential requirements:

---

[6] Loc. cit.
[4] Loc. cit.

(*a*) The presence of the search transmitter does not appreciably affect the deflection of the disc.

(*b*) The output of the transmitter is determined solely by the position of the mouth of the tube in the sound field; i.e., it is independent of the angle at which the tube may be pointing, and of any rotation about its axis.

(*c*) The output of the transmitter with the mouth of the tube closed must be small compared to that with the mouth open.

(*d*) The form of the sound wave must be such that the relation between the velocity and pressure at the mouth of the tube is known.

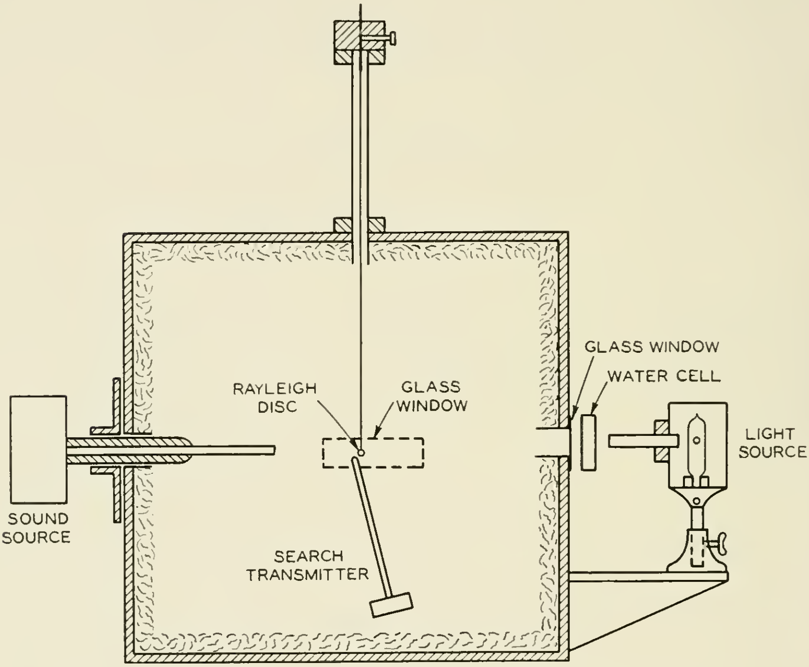


Fig. 18—Search transmitter calibration by means of Rayleigh disc.

Measurement of Pressure in Ear Canal Without Receiver

Pressure measurements are made as far inward in the canal as consistent with perfect safety to and comparative comfort of the subject. Most ears show a distinct bend in the canal accompanied by a flange which is mainly responsible for obstructing the view of the tympanum. It has been found possible to make measurements at a distance of from 0.5 cm. to 1 cm. past that flange, toward the tympanum. It is doubtful if measurements can be made at a point much closer when all circumstances are considered, e.g., the frequency range to be covered and the importance of incurring no risks whatever.

The region inward beyond this bend is apparently one of rather uniform pressure, usually higher than in the outward positions of the canal, so that there is little chance of a measurement being made at the pressure node of a distinct standing wave pattern. This probably is accounted for by the irregularity of the passage and the character of the canal walls.

Receiver Calibration

At each frequency, response measurements were made on eight ears as described, and the average computed. The results are shown in Fig. 15.

When the response-frequency characteristic of one receiver has been determined by the method described, any other receiver may be calibrated by a direct comparison. The response-frequency characteristic on the ear shown in Fig. 16 was obtained in this manner.

Abstracts of Technical Articles from Bell System Sources

A New High Vacuum System.[1] J. A. Becker and E. K. Jaycox. A new high vacuum pumping system is described in which oil is used in a diffusion type pump, and a trap containing activated cocoanut charcoal replaces the usual liquid air trap. The system is capable of attaining a pressure of 2×10^{-8} mm. Hg. A high degree of vacuum can be attained at least as quickly as with a mercury diffusion pump and liquid air trap. The system is especially adapted to maintaining a low pressure for several days in apparatus which cannot conveniently be sealed off from the pumps.

Phenomena in Oxide Coated Filaments II. Origin of Enhanced Emission.[2] J. A. Becker and R. W. Sears. Various theories have been advanced regarding the mechanism of emission of electrons from oxide coated filaments. These theories postulate that: (1) the active layer is (*a*) at the outer oxide surface, (*b*) at the core-oxide interface; (2) the thermionic electrons come from (*a*) the adsorbed barium, (*b*) the oxide just underneath the adsorbed layer; (3) the current is carried through the oxide coating by (*a*) an entirely electrolytic process, (*b*) thermionic electrons which come from the core, diffuse through the pores of the coating and form a space charge therein, (*c*) electrolytic conduction through the oxide crystals and thermionic conduction between crystals, (*d*) electronic conduction, a small portion being carried by ions. A number of experiments were designed to test these various hypotheses. These experiments show that: (1) When barium is brought to the outer surface of the oxide, either by electrolysis or evaporation from an external source, the emission increases at first, passes through a maximum and then decreases. This change in activity is similar to that for barium on tungsten. (2) When oxygen is brought to the surface of the oxide of an activated filament, the activity decreases rapidly at first and then more slowly. (3) In these two respects, a filament with a core made of an alloy called "Konel" and consisting of nickel, cobalt, iron and titanium acts just like filaments with other cores. (4) When the oxide was stripped from a Konel core filament, the activity decreased by a factor of 6000. (5) The emission-limited current is independent of the area of the core provided that the area of the outer oxide surface remains constant. (6) The conductivity of the oxide varies with the time of sending current through the oxide. (7) The conduction current in the oxide obeys Ohm's law and does not saturate even though its value is hundreds of times larger than the saturated emission. (8) The oxide acquires a positive potential with respect to the core regardless of whether the space current is limited by space charge or by emission. This potential varies linearly with the space current drawn to the plate and is of the order of a few tenths of a volt. (9) The emission for the optimum amount of barium on the oxide surface depends upon the previous treatment of the oxide. From these results we conclude that: (1) The active layer is at the outer oxide surface. The activity depends on the concentration of barium and oxygen on this surface and also upon the amount of metallic barium dispersed through the oxide. The core material does not directly affect the emission but it does greatly affect the ease with which free barium is produced by heat treatment or electrolysis. (2) The thermionic electrons originate in the oxide just underneath the adsorbed barium. (3) Most of the current through the oxide is conducted by electrons, a small portion being carried by barium and oxygen ions.

---

[1] *Rev. Sci. Instruments,* December, 1931.

[2] *Phys. Rev.*, December 15, 1931.

Barkhausen Effect: Orientation of Magnetization in Elementary Domains.[3] RICHARD M. BOZORTH. Brief mention is made of previously published work on the nature of the discontinuities in magnetization discovered by Barkhausen in 1919. Including the recent results described in this note, the experimental data now indicate that changes in the magnetization of ferromagnetic materials occur in the following way: The material is composed of small regions or "elementary domains" (of the order of 10^{-8} cm.^{3}), each of which is generally magnetized to saturation in a different direction. As the strength of the applied magnetic field increases, the magnetization in some of the domains changes suddenly from saturation in one direction to saturation in another direction associated with less potential energy. The change in each domain gives rise to a single click in the telephone receiver which terminates the apparatus usually used to observe the effect. In annealed materials in which the crystal grains are much larger than the domains, the direction of magnetization within each domain depends on the orientation of the crystal grain in which it is situated, and coincides with its direction of easy magnetization as determined by separate experiments on large single crystals.

[3] *Phys. Rev.*, January 15, 1932.

The Rapid Record Oscillograph in Sound Picture Studies.[4] A. M. Curtis, T. E. Shea, and C. H. Rumpel. This paper describes a special oscillograph which was designed for making rapid records in sound picture studies. The oscillograph is briefly described, and illustrations are presented of records obtained in making the following studies: microphonic action of vacuum tubes; noise levels in amplifiers; investigations on rectifiers; studies on light valve clash; action of the biasing current of light valves as used in noiseless recording by the variable density method; acoustical studies showing the rise and decay of transients; loud speaker selection with regard to load carrying capacity and mechanical flutter investigations of reproducer sets.

Vertical Sound Records—Recent Fundamental Advances in Mechanical Records on "Wax."[5] H. A. Frederick. This paper describes recent progress which has been made in laboratory studies of mechanical records of sound cut on a wax disk. Both theoretical and experimental investigations indicate that a phonograph record, cut with vertical undulations instead of the more usual lateral undulations possesses fundamental advantages. The principal improvement comes from a marked increase in the volume and frequency range over which faithful reproduction may be obtained. A higher volume level can be recorded for the same groove spacing and speed. More playing time can be provided with a given size of record and volume level since, for these conditions, both the groove spacing and speed may be reduced. Improvements in methods of processing the stampers and in the record material give a large reduction in surface noise and hence a corresponding increase in the volume range. With these improvements the frequency range which can be reproduced satisfactorily can be extended nearly an octave to 8000 to 10,000 cycles. Other improvements incidental to the improvements noted above are great improvement in the quality of reproduction obtainable directly from a soft "wax" record and a great extension in the life of the hard record.

Effect of Shore Station Location Upon Signals.[6] R. A. Heising. Experiments are described for ascertaining the attenuation suffered by the unreflected wave in traversing relatively small amounts of land between the seashore and hypothetical inland sites. The results show 8 to 12 db attenuation for 1 mile inland with greater attenuation thereafter for unfavorable terrain. Swampy ground produces small attenuation. The classical theory of wave transmission past a straight edge used in optics is applied to explain the reduction. Coexisting phenomena are mentioned.

---

[4] *Jour. S. M. P. E.*, January, 1932.
[5] *Jour. S. M. P. E.*, February, 1932.
[6] *Proc. I. R. E.*, January, 1932.

Oxidation Studies of Rubber, Gutta-Percha, and Balata Hydrocarbons.[7] A. R. KEMP, W. S. BISHOP, and P. A. LASSELLE. The oxidation mechanism of rubber and gutta-percha hydrocarbons has been studied. Rubber hydrocarbon in sheet form oxidizes more slowly and less completely than precipitated gutta-percha, which is believed to be due to the smaller surface exposure of the former material. Gutta-percha hydrocarbon in finely divided form oxidizes to a fairly definite degree in oxygen at room temperature, corresponding to a weight increase of about 38 per cent. The length of the autocatalytic induction periods for rubber and gutta-percha varies over a wide range and is shortened by heating the hydrocarbon in high vacuum before oxidation and by exposure to light.

The rate of oxidation of gutta-percha in air, as compared with oxygen, is reduced in proportion to the oxygen concentration, and the induction period is correspondingly increased.

Carbon dioxide, water, formic acid, and formaldehyde are identified in the volatile oxidation products, and their relative amounts determined. Six to eight per cent of the hydrocarbons are converted to volatile oxidation products. The percentage unsaturation of both rubber and gutta-percha hydrocarbons is reduced in proportion to oxygen absorbed. The ratio of hydrogen to carbon decreases as a result of oxidation.

The solid oxidized products are of such a nature that they cannot be resolved into crystalline materials. They are amorphous acid substances, free from aldehyde and ketone groups. They contain a small amount of peroxides; and the acidity, saponification value, and other properties indicate that most of the oxygen is combined in the form of hydroxyl, carboxyl, and lactonic groups. The mechanism of oxidation of rubber and gutta-percha appears to be the same, and the possibility of a chain mechanism to explain the facts is discussed.

Modern Developments in Precision Clocks.[8] A. L. LOOMIS and W. A. MARRISON. A discussion of precision clock requirements is given in terms of the general equations of motion of an oscillator and extended specifically to the gravity pendulum and the crystal oscillator types.

One of the largest sources of error in pendulum clocks is due to variations in amplitude. The amount of this effect is given, and the usual methods for reducing it by keeping the pendulum swing small are discussed. A method which is being used successfully by Mr. Loomis for controlling the amplitude at a fixed value is described.

---

[7] *Ind. and Engg. Chem.*, December, 1931.

[8] Presented at A. I. E. E. Midwinter Convention, Jan. 25–29, 1932, New York, N. Y., as a part of the Symposium on Time and Time Services. To be published in full as *Monograph B-656, Bell Telephone System Technical Publications*, and available upon request to Bell Telephone Laboratories, New York, N. Y.

The effect of atmospheric pressure on the rate is discussed and it is shown that the four chief rate controlling factors involved can be made to annul each other at a critical pressure for a given pendulum.

Factors that affect the length of a pendulum, such as temperature coefficient of the material, aging, etc., and some factors that affect the restoring forces, are discussed since they affect the period directly.

The effect of the phase of applied driving force in an oscillator is also an important factor. In the case of a pendulum the impulse should be delivered at the instant when the velocity is maximum, that is, at the center of the swing. If it is applied earlier, the rate is momentarily increased, and conversely.

A brief description is given of the Shortt clock, which has established an enviable record for timekeeping in some of the world's outstanding time observatories. The installation of three of these clocks in The Loomis Laboratory at Tuxedo Park, N. Y. is described and illustrated.

The crystal clocks used in Bell Telephone Laboratories, N. Y. City, are described briefly, as well as some of the outstanding features of their use. These "oscillators" were built primarily as a precise standard of frequency, but have been found in addition to serve exceedingly well as timekeepers.

Performance data are given for the crystal clocks, for the three Shortt clocks in The Loomis Laboratory, and for clocks in a number of representative national time observatories.

In addition a brief account is given of a continuous comparison which is being made between the crystal clocks in Bell Telephone Laboratories and the Shortt clocks in The Loomis Laboratory.

The Nature of Metals in Relation to their Properties.[9] EARLE E. SCHUMACHER. The methods of extracting metals from their ores and fabricating them constituted the art of metallurgy in the older sense, an art whose development has closely paralleled the rise of civilization. The modern science of metallurgy, on the other hand, concerns itself to a large extent with the explanation as to why metals behave as they do and in particular why the methods employed in the art produce the effects they do. This science, as distinct from the art, is of comparatively recent origin. It is, indeed, only since the development of the tools used in modern research, notably the microscope and x-ray, that important steps have been made toward the analysis of metallic structure. Rapid progress has been made in the last twenty years and today there is available a vast fund of information in regard to the nature of metals in relation to their properties.

---

[9] *Scientific Monthly*, January, 1932.

Two of the most important properties of a metal are its hardness and strength. These properties are closely related and in general the magnitude of one indicates the magnitude of the other. The development of the relationship between the nature of a metal and the hardness and strength constitutes one of the most interesting chapters in the science of modern metallurgy.

Noise Measurement.[10] S. K. WOLF and G. T. STANTON. The instrumental measurement of noise presents difficulties that have in the past generally defeated its successful accomplishment. While noise exists in a physical state and certain of its quantities are susceptible to direct measurement, the magnitude of a noise is evaluated through the interpretation of the human ear. The ear is non-linear in its evaluation of the various factors of noise. The degree and nature of the ear's non-linearity to the principal factors is discussed, with respect to the chief interpretative impression, that of loudness.

Audiometric measurements approached a more proper evaluation of noise, but in addition to dependency upon human judgment, were only approximate, and represented comparisons of physiological effects of noise rather than true noise values.

An instrument is described that measures intensity expressed in terms of loudness, evaluated for frequency and duration, and which combines portions of a complex wave shape in a suitable manner. The characteristics of the meter and the ear are compared. The readings are in decibels above a zero reference point near the threshold of audibility. The selection and meaning of this scale is explained. Where it is desired to analyze the pitch or frequency of a noise, an analyzer attachment permits either band or single-frequency analysis. Some limitations in its use in making noise measurements are discussed.

[10] *Jour. S. M. P. E.*, December, 1931.

Contributors to this Issue

H. A. Affel, S.B. in Electrical Engineering, Massachusetts Institute of Technology, 1914; Research Assistant in Electrical Engineering, 1914–16. Engineering Department and the Department of Development and Research, American Telephone and Telegraph Company, 1916–. Mr. Affel has been engaged chiefly in development work connected with carrier telephone and telegraph systems.

C. R. Avery, B.S. in Mechanical Engineering, University of Michigan, 1918; M.S., 1920. Western Electric Company at Chicago, New York, and Kearny, New Jersey, 1920–. As a development engineer, Mr. Avery has helped establish efficient methods of manufacture for enameled wire, paper and textile insulated wire, and telephone cable.

D. R. Brobst, E.E., Lehigh University, 1917; First Lieutenant, Coast Artillery Corps, U. S. Army, 1917–19; Engineering Department, Western Electric Company, 1922–25; Bell Telephone Laboratories, 1925–. Mr. Brobst's work has been connected with the development of central office wire and cable.

R. W. Chesnut, A.B., Harvard University, 1917; War Department of French Government, 1917; U. S. Army, 1917–19; Engineering Department, Western Electric Company, 1920–25; Bell Telephone Laboratories, 1925–. Mr. Chesnut has been engaged in the development of carrier telephone and long-wave radio systems.

Hal F. Fruth, B.A., M.A., St. John's College, Minnesota, 1914; M.S., University of Minnesota, 1922; Ph.D., University of Michigan, 1926; Research Assistant and Physics Instructor, University of Minnesota, 1920–22; Physics Instructor, University of Michigan, 1922–25; Physicist, Western Electric Company, Development Department, 1925–. Dr. Fruth has been engaged in development of manufacturing processes for mica condensers, microphones, resistances, wire enameling and gas burner design.

W. S. Gorton, A.B., Johns Hopkins University, 1908; M.A., 1910; Ph.D., 1914. Instructor in Electrical Engineering, Massachusetts Institute of Technology, 1911–13; Physicist, Brady Urological Institute, Johns Hopkins Hospital, 1914–17; Associate Physicist, Bureau of Standards, 1917–19; Western Electric Co., 1919–25; Bell Telephone Laboratories, 1925–. Dr. Gorton has worked on voice-operated telephone repeaters and on the development and manufacture of submarine telegraph and telephone cables.

C. H. G. Gray, B.Sc., Massachusetts Institute of Technology, 1918; Instructor in Electrical Engineering, Massachusetts Institute of Technology, 1918–19; Engineering Department, Western Electric Company, 1919–25; Bell Telephone Laboratories, 1925–. Mr. Gray has been engaged in the development and application of machine testing methods for the inspection of transmitters and receivers.

Gordon R. Harris, B.S., Carnegie Institute of Technology, 1927; Bell Telephone Laboratories, 1927–. Mr. Harris has been engaged in the development of filters for carrier telephone and telegraph systems.

A. H. Inglis, B.A., Yale University, 1914; Western Electric Company, 1914–17; American Expeditionary Forces, 1917–19; American Telephone and Telegraph Company, Department of Development and Research, 1919–. Mr. Inglis has been engaged principally in the development of telephone instruments and sets.

R. T. Jenkins, B.S. in Electrical Engineering, Cooper Union, 1917; E.E., 1920. Engineering Department, Western Electric Company, 1916–25; Research Department, Bell Telephone Laboratories, 1925–. Mr. Jenkins' work has been largely in connection with transmission testing methods and apparatus, and with electro-acoustic calibrations.

W. C. Jones, B.S. in E.E., Colorado College, 1913; Western Electric Company, 1913–25; Bell Telephone Laboratories, 1925–. As Transmission Instruments Development Engineer, Mr. Jones has specialized in the development and application of instruments for the transmission of speech and music.

H. Kress, M.E., Staatliche Hoehere Maschinenbauschule, Frankfurt Am Main, 1926. After graduation, Mr. Kress engaged in engineering work in Germany. In 1929, he joined the Western Electric Company as a development engineer and his work has dealt principally with textile insulated wire.

E. B. Wood, B.S., Princeton University, 1915; M.A., Princeton, 1916; Teaching Staff, Cascadilla School, 1916–17; Captain, Coast Artillery Corps, U. S. Army, 1917–19; Teaching Staff, Pratt Institute, 1919–20; Engineering Department, Western Electric Company, 1920–25; Bell Telephone Laboratories, 1925–. Mr. Wood's work has been connected mainly with the development of central office wire and cable and humidity testing equipment.

The Bell System Technical Journal

July, 1932

Henry as an Electrical Pioneer

By BANCROFT GHERARDI

An address presented at the Meeting of the National Academy of Sciences, April 25, 1932, at Washington, D. C. in commemoration of the Hundredth Anniversary of the Electrical Discoveries of Joseph Henry.

The author expresses his appreciation of the researches of Dr. Harold S. Osborne and Mr. A. M. Dowling in establishing many of the facts upon which this address is based, and of other assistance in its preparation. This research developed many facts of interest and its results are separately published as a supplement to this issue of the *Bell System Technical Journal.*

THIS evening we are gathered together to celebrate the hundredth anniversary of the electrical discoveries of Joseph Henry. I have been assigned the honorable and pleasant duty of speaking of Henry's work as an electrical pioneer. According to the Century Dictionary, a pioneer is one who goes before and leads or prepares the way for others coming after, specifically, the first or early explorer or experimenter in any department of human enterprise. Surely no one is better entitled to the honorable title of pioneer than is Joseph Henry.

Let us look a bit at what this country was about 100 years ago. The population of the United States in 1830 was 12,900,000. The largest city in the country, New York, had a population of 203,000. Philadelphia was the next largest city with 80,000. Washington, the capital of the nation, had a population of 18,800, which included 2,330 slaves. There were twenty-four states in the Union and of these the most westerly was Missouri. Chicago, now the second city of the country, was just being laid out near Fort Dearborn and was not organized as a village until 1833. Detroit, now the fourth city, had a population of 2,222 and Cleveland had 1,076.

Highways were largely trails or unpaved roads. Canals played an important part in the transportation system of the country. The first railroad in the United States began operation by steam in 1830. In 1836 trains between Albany and Schenectady were still pulled by horses. The first steamship, equipped with both sails and steam, had crossed the ocean in 1819, but it was not until 1838 that a ship depending primarily upon steam for its propelling power made the transatlantic trip. Communications were solely dependent upon transportation and subject to all of its uncertainties and lack of speed.

Mills and factories were just beginning to be operated by steam power, having previously been dependent upon water power or man power with occasional use of horses or windmills. Lighting was done principally by sperm oil lamps or by candles. Baltimore was the first American city to adopt gas lighting on a large scale in 1817. Gas was introduced in New York City in 1823, but even after 1830 many large cities went for years without a gas system. In 1833 a petition was addressed to the Common Council of Philadelphia protesting against the use of gas "as ignitible as gunpowder and as nearly fatal in its effects as regards the immense destruction of property."[1] There was no commercial application of electricity. At that time electricity was a matter of Leyden jars and of pith balls, of galvanic batteries and of twitching frogs' legs, of crude galvanometers, and of feeble magnets.

Such was the condition of this country when Joseph Henry was Professor of Mathematics and Natural Philosophy at the Albany Academy about 1830.

But "The old order changeth, yielding place to the new," and Joseph Henry was one of those who made major contributions to this change. Just what did Henry do that contributed to the development of the electric art of today? As you have heard from Professor Magie, we may properly claim priority for Henry for the following among his many electrical contributions:

Henry constructed powerful electromagnets by the use of insulated wire on the magnet core and by using more than one layer of winding. (1829 and 1830)

Henry indicated the proper proportioning of magnet windings, external circuit resistance and electrical battery arrangement for effective operation. (1830)

Henry constructed the first motor embodying an electromagnet and a commutator. (1831)

Henry constructed the first telegraph using an electromagnet as the receiving element and demonstrated its operation with a line wire over one mile long. (1831 and 1832)

Henry discovered the property of self-induction of electrical circuits. (1832)

Henry constructed and operated the first electromagnetic relay. (1835)

Henry determined that by the proportioning of the windings of two coils in inductive relationship the voltage in the secondary circuit could be stepped up or stepped down. (1838)

---

[1] "A Popular History of American Invention," W. Kaempffert, 1924, Vol. 1, p. 554.

Henry produced currents in distant circuits by oscillatory discharges and detected them when the two circuits were separated by several hundred feet. (1842)

These contributions by Henry to our electrical knowledge have been essential to the development of practically every commercial application of electricity. Let us now consider some of these.

The idea of the electric telegraph was not new when Henry did his work. It had been considered by others and in January, 1825, Barlow, an English scientist, wrote as follows: "In a very early stage of electromagnetic experiments, it had been suggested, that an instantaneous telegraph might be established by means of conducting wires and compasses. The details of this contrivance are so obvious, and the principles on which it is founded so well understood, that there was only one question which could render the result doubtful, and this was, Is there any diminution of effect by lengthening the conducting wire? . . . I was, therefore, induced to make the trial, but I found such a sensible diminution with only 200 feet of wire, as at once to convince me of the impracticability of the scheme."[2] Even as late as 1837 Wheatstone, another distinguished British scientist, satisfied himself by experiment and convinced others that the development of electromagnetism in soft iron at a distance was impracticable. This was six years after Henry's work on the proportioning of magnet windings and battery arrangement for maximum effect with line wires of substantial resistance had shown him the solution of the telegraph problem, and five years after he himself had constructed and demonstrated his electromagnetic telegraph over a line wire more than a mile long.

Henry's contributions to the telegraph were three in number. He demonstrated that magnetic action and magnetic control could be exercised at considerable distances if the battery and the magnet windings were suitably proportioned and this is the basis of all electric telegraphs of today. He abandoned the galvanometer or compass needle as a receiving device for the electrical impulses and substituted therefor a magnet operating a movable armature, this making possible rapid signaling and audible receiving. This has continued as the basic form of the telegraph circuit, even with the modern printing telegraph systems, in which electric typewriters are controlled by magnets of Henry's type. He constructed and operated the first electromagnetic relay, a device by which the current in the line circuit controls an armature which carries the contact of a local circuit so that the feeble line current, instead of directly controlling the receiving mechanism, merely opens and closes a local circuit and a strong local current performs whatever functions may be necessary in the receiving mechanism.

---

[2] "On the laws of electro-magnetic action," *Edinburgh Philosophical Journal*, Jan., 1825, Vol. XII, p. 105.

These constitute the fundamentals of a complete magnetic telegraph system and left only its perfection in detail, the development of an alphabetical code in dots and dashes, and its commercial exploitation. In 1844 Morse opened his first telegraph line between Baltimore and Washington and in this line utilized the features referred to above, contributed by Henry. From this start the telegraph rapidly grew to a nation-wide and world-wide communication system.

The next important commercial application of electricity came about 30 years later when Bell invented the telephone. As in the case of the telegraph, important use was made of Henry's contributions to the art.

Bell's telephone makes use of Henry's work on magnets, as does every telephone receiver today. The telephone bell, with which each telephone is equipped to give an audible signal so as to attract the attention of the called party, consists essentially of a polarized ringer directly suggested by the receiving device employed by Henry in his first telegraph demonstration. But a telephone by itself, even equipped with a call bell, is merely an interesting scientific toy. The telephone is useful only as there are a number of them at the ends of telephone lines, these lines centering on telephone switchboards for the purposes of interconnection. In all telephone switchboards, whether of the manual or of the automatic type, there are multitudes of relays for the purpose of controlling circuits and signaling apparatus. There are perhaps seventy million telephone relays in the United States today and the prototype of all of these is Henry's electromagnetic relay, dating back to 1835.

Telephony owes still another debt to Henry and to his work. He discovered the characteristic of electrical circuits known as self-induction. This is a property of all electrical circuits unless especially designed to avoid it and self-induction must be taken into account in many phases of electrical design. In long telephone lines it was found to exercise a favorable effect upon telephonic transmission and about 1900 Dr. Michael I. Pupin, a distinguished member of the National Academy of Sciences, showed how self-induction could be added to long telephone lines so as to improve their talking efficiency. Today his invention is very generally used, not only in the lines of this country but throughout the world, and has been an important factor in the extension of long distance telephony and in making possible long telephone cables.

The start of commercial electric light and power systems and of electric traction came shortly after the beginnings of the telephone about 1880. These, likewise, built upon Henry's work. Powerful electromagnets are the basis of every generator and of every electric motor and it would be difficult for any one today to construct either a generator or a motor and avoid the ideas which were contributed by Henry. Commutators which appear first in Henry's motor are essential parts of every direct current generator or motor.

Henry's electric motor was not the first one to be built. Faraday built the first electric motor and it was interesting because it was a continuously rotating device. However, it had no commutator and did not employ electromagnets. Henry's motor was interesting because, while it was a reciprocating device, it was the first to employ a commutator and an electromagnet. A combination of these principles, that is, a rotating motor, electromagnets and commutation gives the basis of the modern electric motor of today.

Henry's comments in 1831 on his electric motor are interesting. "Not much importance, however, is attached to the invention, since the article, in its present state, can only be considered a philosophical toy; although, in the progress of discovery and invention, it is not impossible that the same principle, or some modification of it on a more extended scale, may hereafter be applied to some useful purpose."[3] Much later, in 1876, he writes: "I soon saw, however, that the application of this power was but an indirect method of employing the energy derived from the combustion of coal, and, therefore, could never compete, on the score of expense, with that agent as a means of propelling machinery, but that it might be used in some cases in which expense of power was not a consideration to be weighed against the value of certain objects to be attained."[4] Certainly a prophecy, when we consider today the extent to which this "indirect method of employing the energy derived from the combustion of coal" is utilized because of its convenience for lighting and because of its flexibility for the operation of power units through electric motors.

Modern electrical systems for power and light would be inoperative without the use of auxiliary circuits and equipment for their proper control. In these, extensive use is made of relays and also of other electromagnetic devices using Henry's magnet principle.

While every branch of the electric power art is indebted to Henry, the alternating current system now in such general use and universal for long distance power distribution owes him a peculiar obligation.

[3] "Scientific Writings of Joseph Henry," 1886, Vol. I, p. 54.

[4] "A Memorial of Joseph Henry," 1880, p. 149.

In his work in 1838 he demonstrated that through the proportioning of the windings of two coils in inductive relationship, the voltage in the secondary circuit could be stepped up or stepped down and here we find the genesis of the modern transformer. The transformer, a device without moving parts, is fundamental to every alternating current system, and it is the principal reason why alternating currents are so generally employed today. Through its use, it is possible to design dynamos for operation at the most effective generator voltage and then step up the voltage by means of a transformer to the most efficient level for use on the transmission lines and then at the distant end of these lines by means of other transformers step down the potential to the most efficient and convenient voltage for use on distributing systems.

So fundamental was Henry's work that in it we find contributions even to the youngest child of the electrical family, the radio communication art. Not only does it use his general contributions to the art because of the fact that radio communication is simply a specialized form of telephony or telegraphy employing a particular method of transmission, but Henry's work in detecting the discharges of Leyden jars at a distance of 30 feet and later at distances of several hundred feet certainly foreshadows radio transmission. In 1842 Henry wrote as follows in regard to his experiment in which he observed the inductive effects of a discharge of Leyden jars at a distance of 30 feet: ". . . when it is considered that the magnetism of the needle [his receiving device] is the result of the difference of two actions, it may be further inferred that the diffusion of motion in this case is almost comparable with that of a spark from a flint and steel in the case of light."[5] In the notes made early in 1844 by a student recording Henry's lectures on natural philosophy the following occurs: "Hence the conclusion that every spark of electricity in motion exerts these inductive effects at distances indefinitely great (effects *apparent* at distances of one-half a mile or more); and another ground for the supposition that electricity pervades all space. Each spark sent off from the Electrical Machine in the College Hall sensibly affects the surrounding electricity through the whole village. A fact no more improbable than that light from a candle (probably merely another kind of wave or vibration of the same medium), should produce a sensible effect on the eye at the same distance."[6]

It is certainly a far cry from Henry's observations of inductive effects at distances of a few hundred feet to the present use of radio for broadcasting and for transoceanic telegraph and telephone service, but a consideration of what he himself says about this and of the notes of his student, quoted above, indicates clearly that in his work is the germ of radio transmission.

---

[5] "Scientific Writings of Joseph Henry," 1886, Vol. 1, p. 203.

[6] Notes of Wm. J. Gibson, entitled "Lectures on Natural Philosophy by Professor Henry," Feb. 28, 1844, p. 135.

Had not a fire in the Smithsonian Institution in 1865 destroyed so many of Henry's original records, there is but little doubt that in these records there would have been preserved many other interesting and suggestive things which he did and many significant comments on them. But the record as it is now known to us is sufficient to establish Henry's contributions as outstanding, and to more than justify this distinguished gathering tonight for the purpose of reminding us of our obligation to him.

Before Joseph Henry died in May, 1878, he had seen an extensive commercial application of telegraphy. He had seen the invention of the telephone but not its commercial application. As years have gone by since that date, a greater and greater superstructure of commercial application has been reared upon the foundations laid by Joseph Henry and the other distinguished scientific workers of his time. When writing as early as 1849, he said: "The only reward I ever expected was the consciousness of advancing science, the pleasure of discovering new truths, and the scientific reputation to which these labors would entitle me."[7] This he surely attained. Joseph Henry died with a national and international scientific reputation, Secretary of the Smithsonian Institution, President of the National Academy of Sciences, a friend of Abraham Lincoln and loaded with honors, both at home and abroad. Since then, as time has gone by, it has but added to his greatness, to the esteem in which he is held and to the value of the services which he has rendered to mankind through his work as an electrical pioneer.

---

[7] *Annual Report* of the Smithsonian Institution for 1857, p. 117.

The Cæsium-Oxygen-Silver Photoelectric Cell

An Investigation of the Relations in a Composite Photoelectric Surface

By C. H. PRESCOTT, Jr., and M. J. KELLY

Technique is described permitting the formation of cæsium-oxygen-silver photoelectric cells under controlled conditions. It is shown that the essential conditions are a quantitative control of the degree of oxidation of the silver cathode base and the amount of cæsium generated together with a regulation of the amount of chemical interaction by a control of the time and temperature of the heat treatment.

Variations in sensitivity to integral light at 2,710° K. color temperature are shown as a function of the initial amounts of oxygen and cæsium and the time of heat treatment.

Small amounts of oxygen were permitted to react with the standard cathode surface. The sensitivity of the cathode fell but recovered due to the diffusion of free cæsium to the surface from the underlying material. The effects are shown in relation to the integral sensitivity and the spectral response from 6,000 Å. to 10,000 Å.

The effects of depositing minute amounts of free cæsium upon the standard cathode surface are also shown in relation to the spectral response.

The active surface of the cathode appears to be a film of free cæsium of atomic dimensions adsorbed upon a matrix of cæsium oxide and silver containing free cæsium and a small amount of silver oxide. The spectral characteristics of the photoelectric response appear to depend largely upon the thickness of the surface film of free cæsium. This film thickness is determined by the cæsium concentration in the underlying matrix and is maintained by a diffusion equilibrium.

INTRODUCTION

EARLY studies of the photoelectric effect were made on pure metals, eliminating, in so far as possible, the effects of absorbed gases. But since the alkali metals alone respond appreciably to visible light, and these only to light at the blue end of the spectrum, the development of photoelectric cells of greater response to ordinary light sources has led to the study of thin films of the alkali metals and of various composite surfaces. The enhanced photoelectric activity of the thin films of the alkali metals was first brought out by Ives[1] who also noted that the maximum response and the greatest extension of sensitivity toward the red end of the spectrum were obtained when the film thickness was of the order of one molecular diameter. Later work[2] has shown that the maximum excursion of the photoelectric threshold of an alkali metal film on a metallic base corresponds with the wave-length of the first line of the principal series of the atomic spectrum of the alkali metal.

---

[1] H. E. Ives, *Astrophys. J.*, **60**, 4 (1924).

[2] Ives and Olpin, *Phys. Rev.*, **34**, 117 (1929).

The first composite surface to attain practical importance was the so-called potassium hydride cell discovered by Elster and Geitel[3] in which a potassium surface is sensitized by a glow discharge in hydrogen. This cell has its maximum response at 4,350 Å. and its photoelectric threshold at 5,900 Å. The response of a good potassium hydride (vacuum) cell to a light source at a color temperature of 2,710° K. is about one microampere per lumen, which is eighty times that obtainable with pure potassium surfaces.

This sensitivity is still below that required for many technological applications. Also, the potassium hydride surface is unstable even at ordinary temperatures and may deteriorate rapidly in use or in storage. So there has been a great demand in engineering applications for both a more stable device, and a cell more sensitive to the red and infra-red light which constitutes the major part of the emission from common incandescent light sources.

The photoelectric threshold is a direct measure of the work necessary to liberate an electron from a surface, i.e., the "work function" which also figures in the thermionic effect. That is, both red sensitive photoelectric cells and active thermionic filaments possess low values of the work function. The study of the thermionic effect in adsorbed films of cæsium on tungsten and on oxidized tungsten by Langmuir and Kingdon[4] and by Becker[5] have indicated surprisingly low values of the electron work function. It was to be expected that some similar surface should possess a high order of photoelectric response to red light. Research along these lines has resulted in this and other laboratories in the development of the cæsium-oxygen-silver photoelectric cell. Early work on cells of this type is reported by Koller[6] and by Campbell.[7]

For the cells discussed in this paper the active photoelectric surface is formed on a roughened silver sheet. This is slightly oxidized by making it the cathode in a glow discharge in oxygen. Cæsium is then generated by chemical reaction in a pellet enclosed within the photoelectric cell bulb. Finally, by a proper temperature cycle the cæsium is condensed on the silver oxide surface of the cathode and allowed to react with it to form the active photoelectric surface.

[3] Elster and Geitel, *Phys. Zeit.*, **11**, 257 (1910).

[4] Langmuir and Kingdon, *Phys. Rev.*, **21**, 380 (1923) abstr.
K. H. Kingdon, *Phys. Rev.*, **24**, 510 (1924).
Langmuir and Kingdon, *Proc. Roy. Soc.*, 107-A, 61 (1925).

[5] J. A. Becker, *Phys. Rev.*, **28**, 341 (1926).

[6] L. R. Koller, *Phys. Rev.*, **33**, 1082 (1929) abstr. *Phys. Rev.*, **36**, 1639 (1930).

[7] N. R. Campbell, *Phil. Mag.*, **12**, 173 (1931). "Photoelectric Cells and Their Applications" (Phys. Soc. London, 1930), p. 10.

The cells thus obtained are highly sensitive to red and infra-red light. The maximum response is at 8,000 Å., the response is one third as great at 10,000 Å., and the photoelectric threshold is somewhere in the neighborhood of 12,000 Å. Cells are frequently obtained with a vacuum sensitivity of 35 microamperes per lumen to a light at 2,710° K. color temperature, and their useful life is indefinitely long, even at a temperature of 50° C. But such results obtain only if the cells be produced under controlled and definitely specified conditions. It appears that the product is definitely affected by variations in the quantities of cæsium and oxygen. Also, the course of the chemical reactions (still obscure) and the thickness of the final cæsium thin film are very sensitive to slight variations in the process conditions. In the early stages of this work the results were highly erratic; only occasionally was a useful cell obtained. A large amount of development time has been devoted in these laboratories to the isolation and correlation of the various factors which determine the sensitivity of a finished cell. With the technique now available, cæsium-oxygen-silver photoelectric cells are prepared under conditions of quantity production with sensitivities varying within a factor of two and with a process shrinkage no greater than obtained in the production of high quality thermionic vacuum tubes.

THE STRUCTURAL DETAILS AND METHOD OF PREPARATION OF THE CELLS

The structure of the cells used in this study is shown in Fig. 1. The cathode is a semicylinder of silver 99.9 per cent pure. The anode is a nickel wire mounted in the axis of the cylinder. These are mounted on a stem which is sealed into a spherical bulb of soda-lime glass. Around the stem is suspended an open ring of heavy copper wire between the ends of which is crimped a tube rolled from thin sheet molybdenum which carries the cæsium pellet. This structure is adapted to the initiation of the chemical reaction by induced high-frequency currents with a minimum heating of the oxidized cathode. A nickel shield is placed between the pellet sheath and the cathode to protect the latter from radiation and to deflect the hot cæsium vapor evolved by the pellet.

The essential steps in the process of formation of the active cathode surface are:

1. Formation of a silver oxide film on surface of cathode.
2. Preparation of cæsium.
3. Combination of cæsium with silver oxide surface.

1. *Formation of a Silver Oxide Film on Surface of Cathode*

The cells are sealed to a high vacuum exhaust system and baked out at 400° C. The next step is the oxidation of the surface. It was found that oxidation by a glow discharge in oxygen was the most suitable method. This oxidation could not be done quantitatively and uniformly over the front surface of the cathode because of surface irregularities in the silver and because of variations in the physical conditions of the surface. Means were therefore sought to obtain a uniform surface and one that would be the same in all cells, independent of irregularities that could not be avoided, which were due to past history.

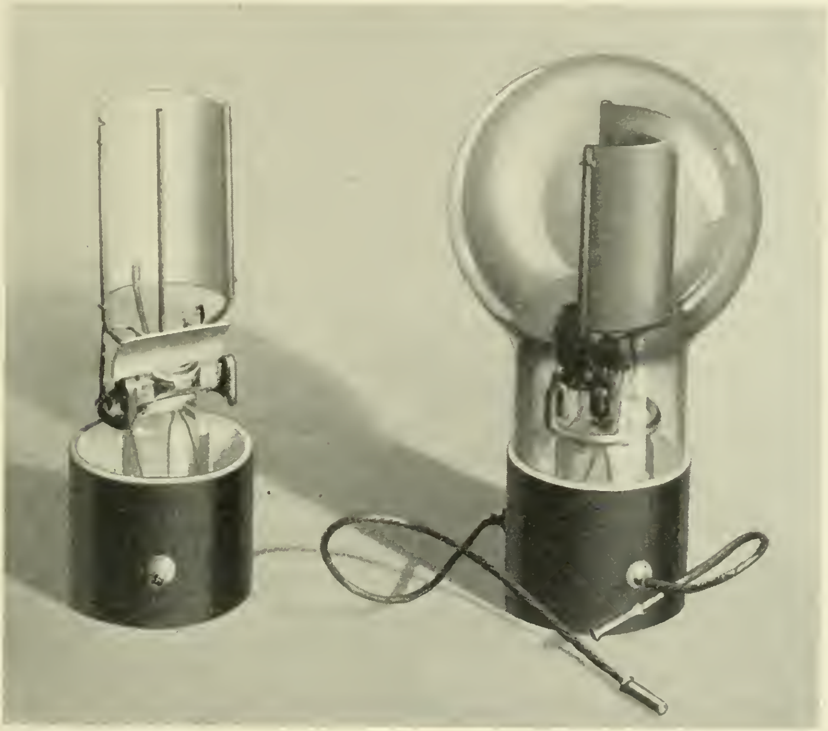


Fig. 1—Western Electric 3A photoelectric cell.

To accomplish this, electrolytic oxygen is admitted to the exhaust system and the cathode is oxidized by a glow discharge until completely black. Then the circuit is held closed while the cathode is heated by ion bombardment until the silver oxide is decomposed, leaving a bright matte surface. After cooling, this process is repeated.

It is repeated a sufficient number of times to establish a cathode surface that is uniformly rough over its entire front surface and is substantially the same for all cells. The exhaust system is then evacuated and pure oxygen admitted to the optimum pressure for quantitative oxidation.

The quantitative oxidation is accomplished by charging a condenser of known capacity to a definite voltage and then discharging it through the cell in series with a suitable resistance. The discharge is in the proper direction for oxidation of the cathode. A double contact telegraph key is used for charge and discharge of the condenser. The number of taps given to the key is then a quantitative measure of the oxidation of the cathode. After the requisite oxidation of each cathode, the system is again evacuated.

As the oxidation proceeds, the surface goes through characteristic changes in color. At 25 "taps" it is yellow, at 50 red, at 75 blue and at 100 a greenish yellow. If the oxidation is continued, the color goes through another cycle becoming a golden yellow, a deep rose red and an olive green. From then on it is quite dark, but under strong light shows several alternations of red and green before ending in black. As will be seen later an oxide thickness corresponding to about 100 taps is the most suitable for use.

2. *Preparation of Cæsium*

The early work demonstrated the necessity of a close control of the amount of cæsium made available for the cathode in each cell. After an examination of a number of chemical systems for preparation of cæsium by a high temperature chemical reaction, a pellet composed of a mixture of cæsium chromate, chromic oxide and powdered aluminum was adopted. These materials are thoroughly mixed in quantitative proportions and ground in an agate mortar. A suitable weight of the mixture is compressed in a die. A slight but uniform loss of material is entailed in the pellet making process. Any desired quantity of pellets differing in weight by not more than ten per cent can be made by this process. This pellet is placed in the molybdenum housing described above and after the quantitative oxidation of the surface and removal of excess oxygen the pellet is heated by high-frequency induction to its kindling temperature.

Sufficient aluminum is supplied to completely reduce the cæsium chromate and chromic oxide. Besides permitting a pellet of convenient size, the chromic oxide is instrumental in furnishing, by its reduction, a large amount of heat. The high-frequency heating serves only to initiate the reaction. The exothermic reaction involves a great rise in temperature and causes the immediate and complete expulsion of all cæsium. This allows the high-frequency heating with the attendant hazard of heating the cathode and reducing a part of the silver oxide to be a minimum. The cæsium travels in straight lines from the pellet housing and is condensed on the glass wall of the bulb. A shield above the housing prevents any cæsium, as it is expelled from the pellet, from impinging on the cathode surface.

3. *Transfer of Cæsium to the Cathode Surface*

The transfer of the cæsium to the cathode surface and its reaction with the silver oxide film to the proper extent is probably the most difficult process to control. The ideal process requires the transfer of the cæsium from the glass wall of the envelope to the cathode surface without the reduction of any of the silver oxide by the heat required to bring about this cæsium transfer, the reaction of the silver oxide with the greater portion of the cæsium and the leaving of a sufficient amount of uncombined cæsium to supply the required volume concentration of cæsium and to cover the entire cathode surface with an equilibrium thin film of cæsium. The practical difficulties in carrying out this process are due to the fact that, at the temperatures required to transfer in a reasonably short time the cæsium from the glass wall to the cathode, the silver oxide has an appreciable rate of decomposition. This difficulty was overcome so far as is possible by heating the glass wall selectively at a controlled temperature. There is a material lag in the cathode temperature in this process. Thus the cathode is kept at as low a temperature as possible while the cæsium is transferred to it at a sufficiently rapid rate.

This is accomplished by surrounding the cell with a stream of hot air for approximately 30 minutes. A glass chimney is placed around the cell. The chimney fits into a transite manifold containing heating coils through which compressed air is fed at the rate of 0.5 liter per second per chimney. The temperature of the air stream is controlled to within 5° C.

If this process is carried on in the usual electric oven where a portion of the winding is exposed, the cathode is heated preferentially due to the radiant heat and it is almost impossible to obtain an active surface due to silver oxide reduction by temperature, as well as to the fact that with the cathode the hottest surface, the cæsium will tend to condense elsewhere. Even with all windings of the oven covered by asbestos sheet so that substantially all heat is due to convection, the process is infinitely more difficult of control than with the hot air stream oven.

It was not found practicable, even with such exact control of the silver oxide film, the amount of cæsium and of the temperatures of the photo-cell parts, to give the cell a definite time-temperature cycle in the transfer process. It is necessary to follow the growth of the photo and thermionic currents as the heating continues. A small tungsten lamp is mounted in a fixed position with respect to the cell and after the heating has progressed for some 20 minutes, observations are made each minute on the total thermionic and photoelectric currents. These currents rise in value as the surface is built up and, by experience, a definite point on the growth curve is found where the hot air stream should be discontinued in order to obtain the optimum surface.

After the cell has cooled it is sealed from the exhaust system if a vacuum cell is desired, or filled to the required pressure with an inert gas, usually argon, if a gass filled cell is desired.

Surface Structure

In the study of the photosensitive cathode surface some information regarding the physical structure and chemical nature of the surface has been gained by direct examination and analysis. The physical nature of the roughened silver may be seen from microphotographs of the cathode taken following the roughening oxidation and reduction which precedes the quantitative oxidation.

Fig. 2*a* is a plan view taken at 1,530 diameters magnification. Fig. 2*b* is a transverse section of the cathode supported in a heavy nickel plate taken at 200 diameters. Fig. 2*c* is a detail view of the front surface of the transverse section taken at 2,450 diameters. It would appear that the effect of the oxidation and reduction is to etch out the polished silver sheet, giving the surface elements a random orientation but not causing any great increase in surface area. If we judge the length of a line element to be doubled, this indicates a four-fold increase in area.

To determine the amount of oxygen entering into combination at each discharge of the condenser, three cells were given 50, 75 and 100 taps respectively of the key controlling the condenser discharge. The excess oxygen was removed and the exhaust manifold trapped off by a mercury seal. Each cathode in turn was then heated by induced high-frequency current till the silver oxide decomposed, and the pressure developed in the exhaust system was measured. The pressures and amounts of oxygen as microgram-molecules of O_2 are given in Table 1.

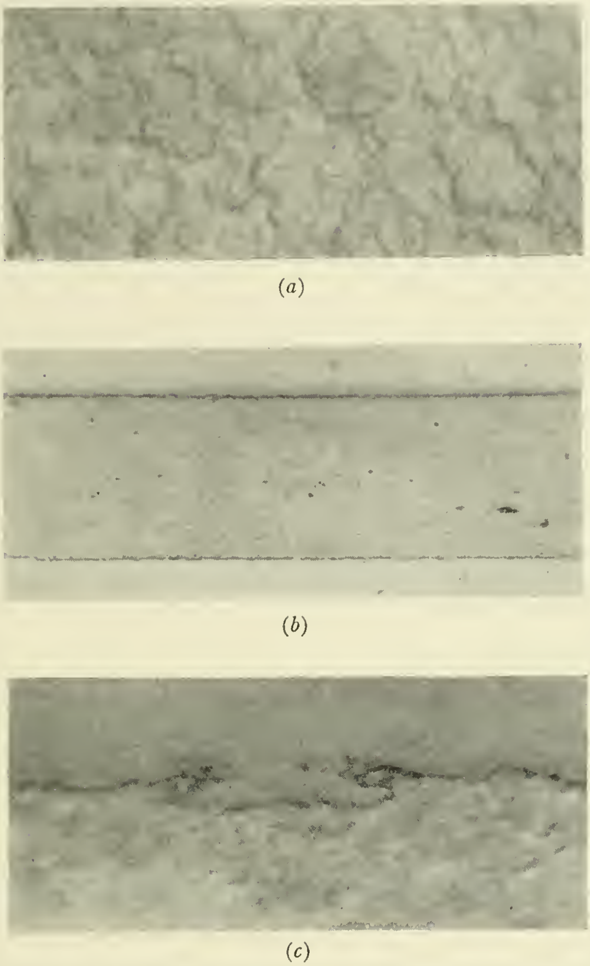


Fig. 2—Microphotographs of roughened silver cathode base. (*a*) Plan view, 1530 diameters. (*b*) Transverse section, 200 diameters. (*c*) Transverse section, 2450 diameters.

TABLE 1

No. of Taps	Pressure	Micro-moles of O_2	Micro-moles per 100 Taps
50	0.0179 mm.	2.24	4.48
75	0.0256	3.20	4.27
100	0.0369	4.62	4.62
Weighted average			4.45

This value of 4.45 micro-moles of oxygen per 100 taps corresponds to 1.32 molecules of O_2 or 2.64 atoms per electron equivalent of the charge transported in the glow discharge, assuming complete discharge of the condenser. This value entails a high efficiency in the combination of positive ions with the surface, and also predominantly polyatomic ions at least as complex as O_3^+.

Microchemical analyses were performed on several cells to determine the cathode composition and the final distribution of the cæsium. With an accuracy of about 3 per cent all the initial cæsium was recovered, 68 per cent being recovered from the silver cathode and the rest from the inner surface of the glass bulb. No cæsium was found in the pellet residue. In the cathode surface was also found an average of 0.13 milligram of undecomposed silver oxide. But the actual amount may be greater for we have later evidence that at least 7 per cent of the cæsium on the cathode occurs as the free metal which would reduce silver oxide when the cathode is extracted with water and dilute acid prior to the analysis.

The optimum conditions, as shown later, are obtained with 4.1 milligrams of cæsium chromate in the pellet and a "ratio" of 19 condenser discharges per milligram of cæsium chromate. These are equivalent to 83 micrograms of cæsium and 3.2 micrograms of oxygen (using the factor from Table 1) per square centimeter of the cathode and a ratio of 3.1 atoms of cæsium per atom of oxygen. With 68 per cent retention of the cæsium in the cathode, we obtain 56 micrograms of cæsium per square centimeter and an atomic ratio of 2.1. If we neglect the free cæsium and residual silver oxide in the cathode surface and assume that no oxygen has been lost from the cathode, which implies that the cæsium on the bulb has been oxidized by reaction with water or other constituents of the glass—the atomic ratio of 2.1 suggests that Cs_2O is the main cæsium constituent of the cathode surface. But however probable it may seem, it is not possible to demonstrate this by ordinary analytical means in the presence of free cæsium and residual silver oxide.

These initial and retained amounts of cæsium are equivalent respectively to 910 and 620 atomic layers on the plane cathode surface. If we assume the surface to be increased by a factor of four in roughening, we find the cæsium retained in the cathode surface to be equivalent to roughly 155 atomic layers of free cæsium. This cæsium will, of course, occur mainly as cæsium oxide and be mixed with the finely divided silver from the reduced silver oxide. It would comprise about 50 layers of Cs_2O molecules.

L. R. Koller[6] has given quantitative data on surfaces which were

[6] Loc. cit.

formed by admitting oxygen to a cathode surface covered with a thick adsorbed layer of metallic cæsium. He finds a maximum activity at a cæsium-oxygen weight ratio of 110, which corresponds to 13.2 atoms of cæsium per atom of oxygen. This high value may be due in part to an incomplete yield from his cæsium pellets, but we doubt that the surface structure so obtained can be closely compared with that which we have described. The cells described elsewhere in his paper have much more closely analogous surfaces.

N. R. Campbell[7] has performed experiments in which cæsium is diffused slowly into a cell containing an oxidized silver cathode, all held at 184° C. in an aniline bath. He finds the completion of the reaction corresponds to the formation of Cs_2O, or an atomic ratio of 2.

Kingdon and Thompson[8] also report their cathode surfaces, presumably the same as in the cells described by Koller, to consist of Cs_2O.

A Survey of Macroscope Parameters

With uncertainty as to the microscopic physical structure of this surface, and no adequate theoretical basis for the correlation of photoelectric response to such a detailed structure, a macroscopic frame of reference has been essential to the correlation of data. This has been particularly important in the development phases of the project so that the results obtained should be self-sufficient and the systematic investigation of substances and processes should not be conditioned by the anticipated interpretations. From a thermodynamic standpoint, the surface may be considered as a two-dimensional phase or system with two variable components, cæsium and oxygen. If this system were in equilibrium, its properties would depend only on its temperature and the amounts of the components. Since it is far from equilibrium the specification of its condition or "state" requires also its past history, the most vital part of which is the heat treatment.

As the activity of a completed cell is definitely affected by each of the factors controlled in the quantitative technique, it will be seen that the state of the cathode surface is a function of (at least) five parameters, viz: the amount of cæsium, the amount of oxygen, the surface roughness of the silver, and the temperature and time of the heat treatment. Some qualitative experience seems to indicate that variation in surface roughness shows itself chiefly by affecting the efficiency of the glow discharge in the quantitative deposition of oxygen, and to a limited extent involves a greater or less extension of the surface. Also, the temperature and time of heat treatment are to a limited extent compensatory. So a more fundamental specification of the active surface would involve the surface concentrations of cæsium and oxygen, and a degree of interaction which is commensurate with the heating time.

---

[7] Loc. cit.

[8] Kingdon and Thompson, Physics, **1**, 343 (1931).

The best known thin film phenomena in the past have been of the type of cæsium adsorbed on tungsten, a single component film under essentially equilibrium conditions. At least the film is reversibly adsorbed for it may be condensed or evaporated at will by varying the tungsten temperature or the cæsium vapor pressure. Cæsium on oxidized tungsten is again a two-component film. The cæsium is reversibly adsorbed and, after the cæsium is removed, it appears that the adsorption of the oxygen is reversible at a considerably higher temperature.

Langmuir and Villars [9] have published curves showing the thermionic activity of tungsten filaments with varying amounts of adsorbed cæsium and oxygen. Due to the lack of an independent measure of the amount of adsorbed oxygen the data have been presented from a different point of view, resulting in a determination of the heat of adsorption of the oxygen. But we may point out the obvious though tacit assumption that the thermionic and other properties of the surface are uniquely determined in terms of the amounts of adsorbed oxygen and cæsium (and the temperature). This is equivalent to the presumption that the filament is under equilibrium conditions.

In contrast, the two-component film of cæsium and oxygen on silver which comprises the photoelectric cathode is formed by essentially irreversible processes. No cæsium atom which has combined with oxygen can be released, and no oxygen atom split off from the silver surface may be recombined. And the reaction is stopped long before any equilibrium is reached.

To obtain definite knowledge of the relations in the neighborhood of the optimum conditions we have prepared cathode surfaces under carefully controlled conditions. Two different amounts of cæsium were used and the oxygen-cæsium ratio (the number of "taps" per milligram of cæsium chromate) and the time of hot air heating were systematically varied. The first series of cells was made with pellets containing approximately 5 milligrams of cæsium chromate. The oxygen-cæsium ratio was varied from 15 to 30, and three groups heated for 15, 30 and 60 minutes respectively at 220–225° C. In each case, as shown in Fig. 3, the activity as measured with a light source at a color temperature of 2,710° K. goes through a maximum

[9] Langmuir and Villars, *J. Amer. Chem. Soc.*, 53, 486 (1931).

in the neighborhood of the 20 ratio. Also, the maximum activity is significantly higher for the 30-minute heat treatment than for either 15 or 60 minutes. The second series of cells was made with pellets containing 3 milligrams of cæsium chromate. The oxygen-cæsium ratio was varied from 10 to 30 and three groups baked for 7.5, 15 and 30 minutes, respectively. In this series the maximum activity occurs at a ratio of 17 and with a heat treatment of 15 minutes as shown in Fig. 4. One additional point of interest is that with the light pellets the cells have still fair activities at the high ratios for the short heat treatments but are destroyed by the 30-minute heating. This is most probably due to the oxidation of free cæsium by residual silver oxide.

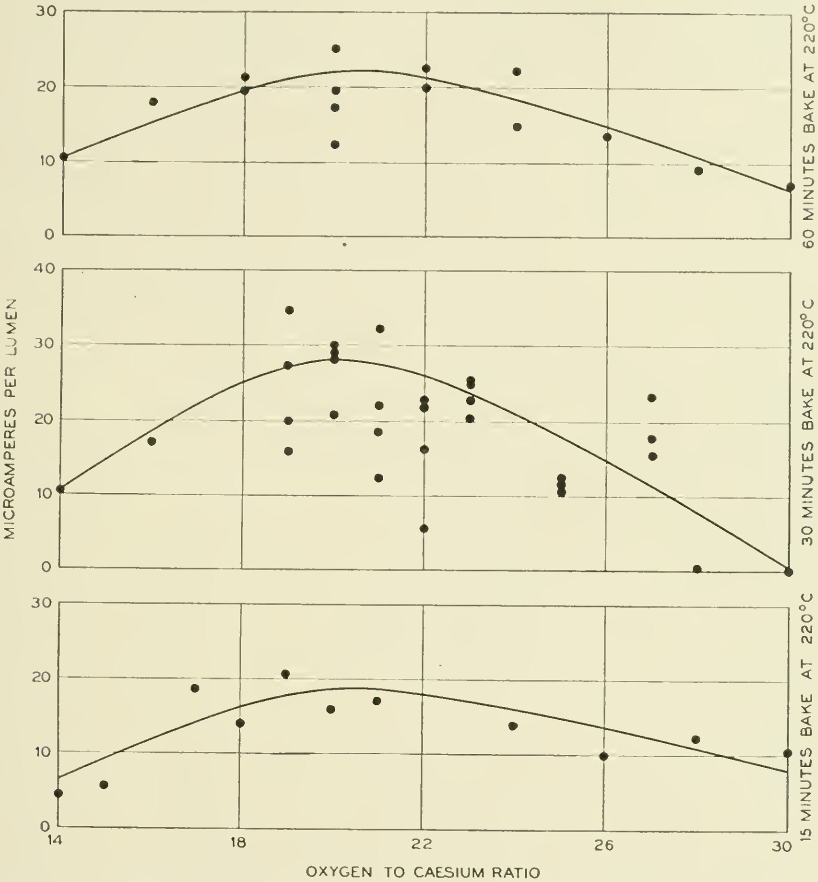


Fig. 3—Integral sensitivity as a function of oxygen-cæsium ratio for 5 mg. pellets.

It is also significant to compare the coordinate effects of both oxygen-cæsium ratio and weight of cæsium chromate upon the activity. To this end in Figs. 5 and 6 for the 15 and 30-minute heat treatments, the ratio is plotted as abscissa, the weight of cæsium chromate as ordinate, and a circle is drawn about each point whose diameter is proportional to the observed activity. It is apparent that the 3-milligram pellets give superior results for the 15-minute heat treatment, and the 5-milligram pellets for the 30-minute treatment. Also the 5-milligram pellets seem to be the better, each considered under its optimum conditions. The 5-milligram pellets also seem to yield cells more stable with respect to excessive heating, cf. Figs. 3 and 4. The pellet weights were in general quite uniform, but due to temporary difficulties the weights of the 5-milligram pellets used in the 30-minute heat treatment scattered considerably as is shown in Fig. 6. Fortunately, this scattering furnishes some detailed evidence as to the relation between activity and pellet weight. The locus of highest activity appears to be an oblique line as expressed in the coordinates of Fig. 6. One might predict that the highest activities would be

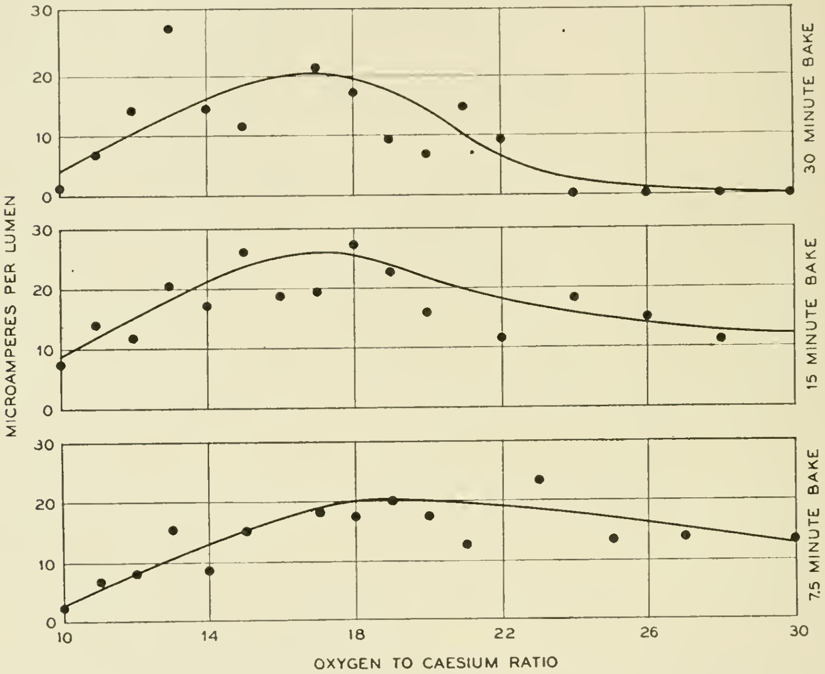


Fig. 4—Integral sensitivity as a function of oxygen-cæsium ratio for 3 mg. pellets.

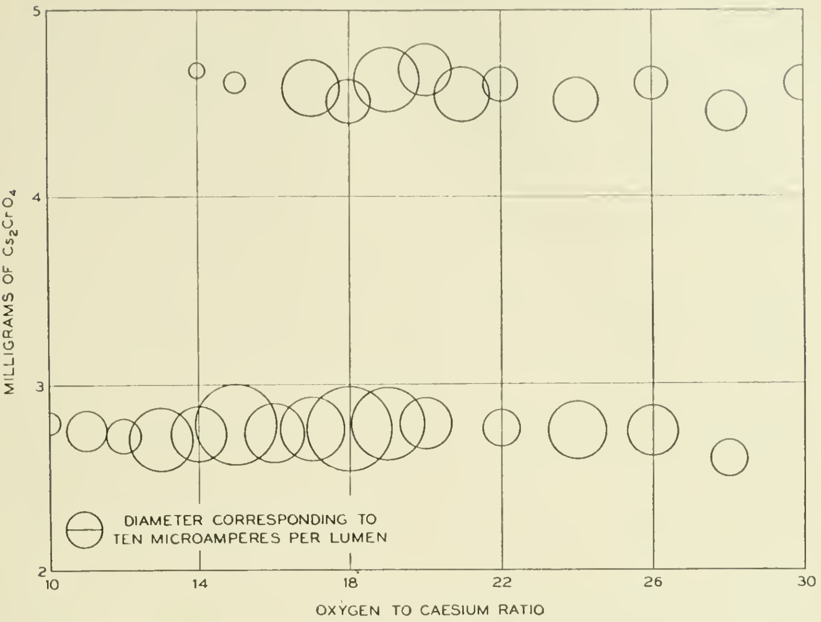


Fig. 5—Integral sensitivity as a function of ratio and weight of cæsium chromate for a 15-minute heat treatment.

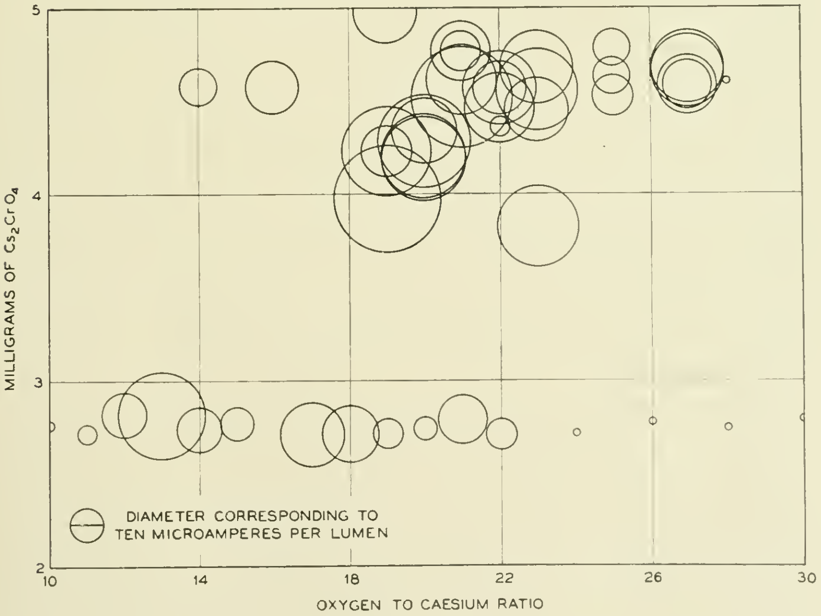


Fig. 6—Integral sensitivity as a function of ratio and weight of cæsium chromate for a 30-minute heat treatment.

obtained by specifying a pellet containing 4.0 to 4.2 milligrams of cæsium chromate, an oxygen-cæsium ratio of 19, and a heat treatment at 220° C. for 25 to 30 minutes. It should be noted, of course, that the above are initial specifications. Two thirds of the cæsium goes to form the matrix and active film of the cathode surface, while the remainder is deposited upon the inner surface of the glass bulb.

Studies of the Cæsium Thin Film

These cells show small fluctuations in sensitivity and shifts in spectral response with time and with the temperature of the surroundings. Even cells prepared under the optimum conditions take several days to stabilize. The integral sensitivity often changes by 5 or 10 per cent, usually to higher values. But further changes occur with changes in temperature. A number of cells aged at 50° C. increased in sensitivity by about 10 per cent in a few days, and then fell to their former stable values after operation for a few days at room temperature. In particular, cells made with large ratios of cæsium to oxygen and short heat treatments, which should be conducive to the presence of small residual amounts of free cæsium, tend to be the least stable, and to decrease rather than gain in integral sensitivity. Such cells are apt also to have low insulation resistance which tends to be unstable in value as compared to the normal cell which has a high and relatively stable insulation resistance.

Since the changes in sensitivity at 50° C. or below are generally reversible, they can not be due to chemical reactions of the irreversible type by which the cells are prepared. These changes seem most easily explicable as due to changes in thickness of a thin film of free cæsium reversibly adsorbed upon the matrix of gross material, many molecules thick, which has been formed upon the silver cathode surface. According to this view it should be possible to correlate the observed fluctuations in sensitivity with variations in film thickness as the free cæsium diffuses in and out of the underlying matrix, or evaporates and recondenses upon the various surfaces of the photoelectric cell interior.

Now it is impossible to differentiate between cæsium adsorbed on the surface, absorbed in the matrix, or in chemical combination, by the type of quantitative data considered in the previous section, for all are placed simultaneously in their appropriate places and in proper amounts by the processes described. Neither is it possible to build up a matrix alone and subsequently place the film upon it with any assurance that the final state shall be comparable to that found in the cells already described. So resort has been had to a variational method of studying the surface film.

We have wished to obtain information regarding the amount of cæsium in the surface film, the amount of free cæsium in the underlying body of the matrix, to check up on the possibility of the diffusion of free cæsium between the surface and body of the matrix and to study the relation between film thickness and the spectral response of the cathode surface. To this end we conducted one series of experiments in which known quantities of oxygen were admitted to cathode surfaces prepared in the standard fashion, following the course of recovery in response with time at various temperatures. In further experiments cæsium vapor was allowed to deposit continuously upon the active surfaces of several cathodes of the same type, while observations were made of the resulting changes in spectral response.

Special cells were made and pumped under as nearly as possible the optimum conditions. The first series of cells were similar in structure to that shown in Fig. 1, except that to the upper end of the bulb was sealed a short glass tube. Within this tube and attached to its upper end was a thin-walled glass bulb containing oxygen at reduced pressure. A steel ball was also included within the tube which could be made to break the glass bulb of oxygen by agitation of the cell. The second series of cells were constructed as shown in Fig. 7. In each cell was supported a glass cylinder closed except for a small orifice pointing obliquely towards the cathode surface. Within this cylinder were a thin-walled bulb of soda-lime glass containing free cæsium and a steel ball with which the bulb could be broken as before.

Monochromatic light was obtained from a Bausch and Lomb No. 2700 glass spectrometer fitted with two slits and calibrated by the makers to 10,000 Å. The light source was a lamp with a helical tungsten filament operated at a color temperature of 2,710° K. Readings were corrected to an equal energy scale using the relative energy curve for tungsten and the dispersion curve of the spectrometer. At the blue end of the spectrum readings were very small due to the low intensity of the light, the low response of the cell, and the wide dispersion of the spectrometer. The values presented are from 6,000 Å to 10,000 Å. The measurements were made with 90 volts across the cells and the photoelectric currents were amplified on a vacuum tube bridge circuit so as to be rapidly readable on a microammeter.

In Table 2 are presented data from four cells to which oxygen was admitted, giving the amounts of oxygen, the equivalent number of atomic layers of cæsium oxidized, the various aging and heat treatments, and the integral sensitivities obtained after the several treatments. In the second row of Table 2 the equivalent layers of cæsium are computed as upon a smooth cathode surface. To obtain actual layers upon the roughened cathode these values must be divided by a factor which we have judged to be approximately four. In addition, since the oxygen flowed freely into the whole structure of the cell, we must consider a film (probably monatomic) of free cæsium adsorbed on the inner surface of the glass bulb which has about four times the area of the cathode. This film must be present since the bulb itself is appreciably photosensitive. In cell *A* the oxygen, calculated as

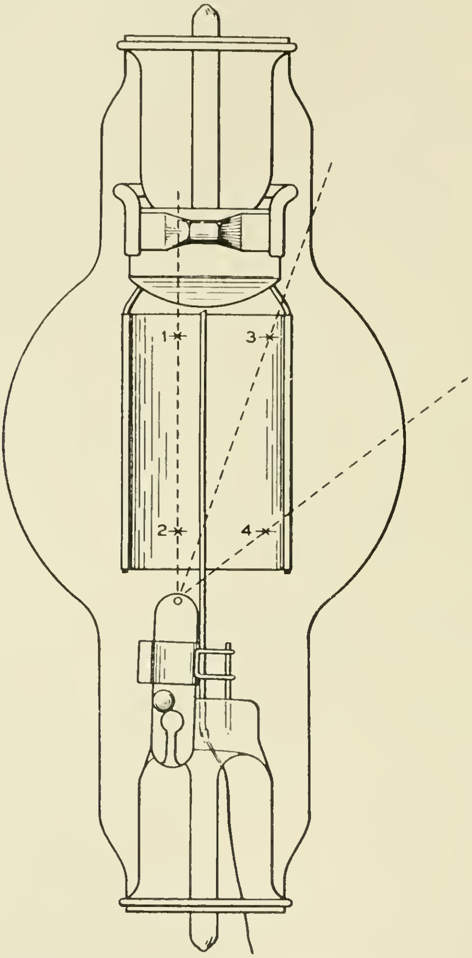


Fig. 7—Cell design for the deposition of cæsium on a cathode surface.

TABLE 2

	Cell *A*		Cell *B*		Cell *C*		Cell *D*	
Moles of O_2	$2.36 \cdot 10^{-8}$		$5.27 \cdot 10^{-8}$		$17.5 \cdot 10^{-8}$		$29.2 \cdot 10^{-8}$	
Equivalent layers of cæsium on smooth cathode	4.01		8.95		29.7		49.7	
Estimated layers of free cæsium oxidized on roughened cathode	0.5		1		6		11	
	Time of Treatment	Integral Sensitivity	Time of Treatment	Integral Sensitivity	Time of Treatment	Integral Sensitivity	Time of Treatment	Integral Sensitivity
Initial sensitivity		34.7 $\frac{\mu \text{ amps.}}{\text{lumen}}$		25.4		26.4		29.1
After releasing oxygen		16.2		4.20		0.25		0.10
After aging at 22° C.	48 hours	20.3	68 hours	6.83			3 hours	0.41
After aging at 75° C.	1.7	28.8	8.7	14.3	0.5 hours	1.22	1.0	0.91
After further aging at 75° C.	3.0	30.8	5.0	15.3	11.5	0.93	5.0	0.82
After aging at 200° C.			5 minutes	13.1	5 minutes	23.4	5 minutes	1.00
After further aging at 200° C.			10	21.1			5	1.00
							10	19.9

equivalent to four layers of cæsium upon a smooth cathode, would be sufficient to oxidize either the monatomic film of cæsium on the bulb, or one atomic layer upon the cathode. We have therefore supposed that about half an atomic layer upon the surface of the cathode was actually oxidized. With the other three cells we have subtracted four "layers" for the bulb and divided the remainder by four to obtain a rough estimate of the actual number of atomic layers of free cæsium upon the cathode surface which were oxidized. These values are shown in the third row of Table 2.

Each cell was first baked for 30 minutes at 75° C. to insure that the surface was in substantial equilibrium. No integral sensitivity changed by more than 10 per cent, and changes in the distribution of spectral response were inappreciable. These results are in agreement with our general experience that these cathodes are stable at 75° C.

All four cells were radically affected by the introduction of oxygen, but none were made entirely inactive. All recovered somewhat at room temperature but the rate of recovery slowed down rapidly. After the recovery had flattened out at room temperature the cells were baked at 75° C. The process of recovery was much accelerated so that cell *A* recovered most of its initial sensitivity, but cells *B*, *C* and *D* did not recover a high sensitivity until after a short heat treatment in an oven at 200° C. Cell *A* losing half a layer of cæsium upon the introduction of oxygen fell to 47 per cent of its initial sensitivity. It recovered to 59 per cent after 48 hours at room temperature, and to 89 per cent after 3 hours at 75° C. Cell *B* losing one layer fell to 16.5 per cent, recovered to 27 per cent after 68 hours at room temperature, to 60 per cent after 14 hours at 75° C., and to 83 per cent after 15 minutes total at 200° C. Cell *C* losing six layers fell to 0.95 per cent, recovered to 3.5 per cent after 12 hours at 75° C., and to 89 per cent after 5 minutes at 200° C. Cell *D* losing eleven layers fell to 0.3 per cent, recovered to 1.4 per cent after 7 hours at room temperature, to 2.8 per cent after 6 hours at 75° C., and to 68 per cent after 20 minutes total at 200° C.

The heat treatment at 200° C. was kept short as reaction between free cæsium and silver oxide and evaporation of cæsium oxide both occur at this temperature and in time would destroy even a normal cathode surface. In the five-minute heat treatments the maximum cathode temperature was doubtless considerably less than 200° C. and the increased effects of the ten-minute treatments which followed are probably due more to the enhanced cathode temperature than to the longer time. The observed recovery must be due to the migration of free cæsium already present as there is no chemical reaction possible under these conditions by which free cæsium could be obtained from the cæsium oxide present.

This recovery of sensitivity has every appearance of accompanying a diffusion process by which cæsium in the matrix is brought to the surface of the cathode. Initially the concentration is uniform throughout the matrix. As soon as the oxygen is released it oxidizes the cæsium on and near the surface until the oxygen is exhausted. The cæsium oxide formed is identical with that already present. The resulting condition will be a matrix with a very sharp gradient from outer surface inward in cæsium concentration. This will cause a rapid initial rate of outward diffusion. But as soon as diffusion occurs the concentration gradient is decreased and the rate of transfer is less rapid since the cæsium is transported over a constantly greater distance under a decreasing diffusion pressure. As the temperature is raised the mobility of the cæsium is greatly increased but the essential nature of the process is the same.

It does not appear that any conclusion can be drawn from these data as to the thickness of the surface film of cæsium. Even in cell *D* with oxygen equivalent to eleven atomic layers of cæsium there was still a finite residual activity immediately after the oxygen reacted, which was not far different in value from that obtained with the six equivalent layers in cell *C*. It seems probable that the matrix is sufficiently spongy that there is a rapid diffusion of oxygen into it and immediate interaction of oxygen with the absorbed cæsium, preventing a complete clean-up of all cæsium on the surface with the amounts of oxygen used. Cell *D* is also noteworthy in that the oxygen used was equivalent to 7 per cent of the total cæsium found on such a cathode. And after the oxidation of this amount of free cæsium it was still possible to develop from this surface a high sensitivity equal to two thirds that of the initial state. So the initial amount of free cæsium must be still greater than 7 per cent of the total cæsium on the surface.

But it is not necessary to suppose a much greater amount since it is known that monomolecular films have a great tendency to form. This may resolve the difficulty in understanding why, after the equilibrium is reestablished, so large a change in volume concentration involves so slight a change in the surface condition. While the thickness of a surface film in equilibrium with a gas or vapor is a definite function of the pressure, the film thickness changes much less rapidly than the pressure when the film is in the neighborhood of one molecule thick. The same type of functional relationship may also obtain when the surface concentration is determined as a function of an underlying volume concentration. Langmuir [10] has shown by the use of the Gibbs adsorption isotherm that in the case of solutions where the solute greatly lowers the surface tension, there is a layer of solute at the surface which is approximately a monomolecular film under a wide range of concentrations in the body of the solution. The same type of relationship may well obtain in the system under discussion, and the concept of a surface film merges into that of a limiting concentration at the surface of the matrix.

Measurements of spectral response were taken with the light incident upon four different positions on the cathode surface. The relative positions are the same as shown in Fig. 7. The effects were quite similar for positions 1 and 3 and again for positions 2 and 4, so curves are shown only for positions 1 and 2 for cells A and B to illustrate the effects at opposite ends of the cathode. In cells C and D the results were so similar for all four positions that curves are shown only for position 1.

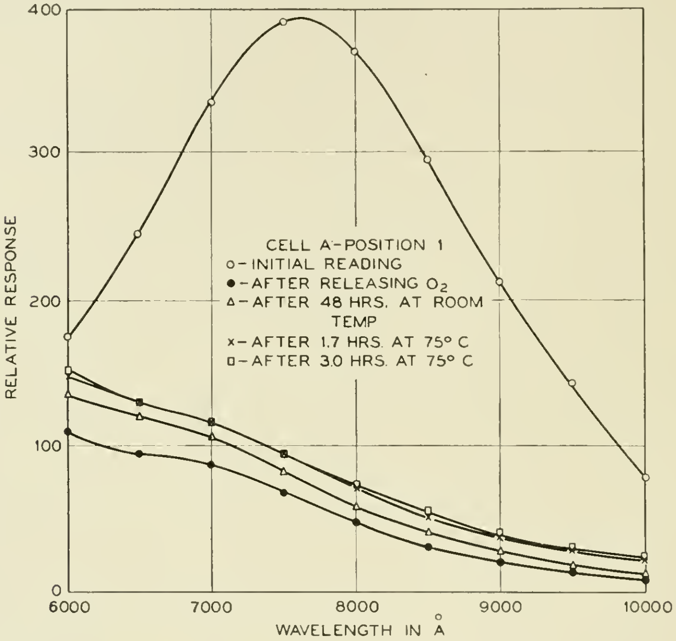


Fig. 8—Relative spectral response to equal energy for cell A, position 1.

In Figs. 8 and 9 are shown the effects upon cell A of the admission of oxygen equivalent to half an atomic layer of free cæsium. Position

---

[10] Irving Langmuir, *Proc. Nat. Acad. Sci.*, 3, 251 (1917).

1 was nearer the end of the bulb at which the oxygen was introduced and was apparently more oxidized than was position 2. The characteristic spectral maximum for position 1, Fig. 8, is completely suppressed and has not reappeared after 4.7 hours of baking at 75° C. At position 2, Fig. 9, the height of the spectral maximum is greatly reduced but recovers considerably on baking, moving towards the longer wave-lengths, and accompanied by an increase in response at 10,000 Å. till it exceeds that of the initial state. In this final state the integral sensitivity of the cell as a whole had nearly recovered its initial value. The large difference in behavior of the two ends of the cathode is doubtless due to the fact that the oxygen was insufficient to oxidize even the superficial layer of free cæsium and reacted wherever it first struck the cathode. The difference between opposite ends of the cathode is less pronounced in the other cells treated with more oxygen.

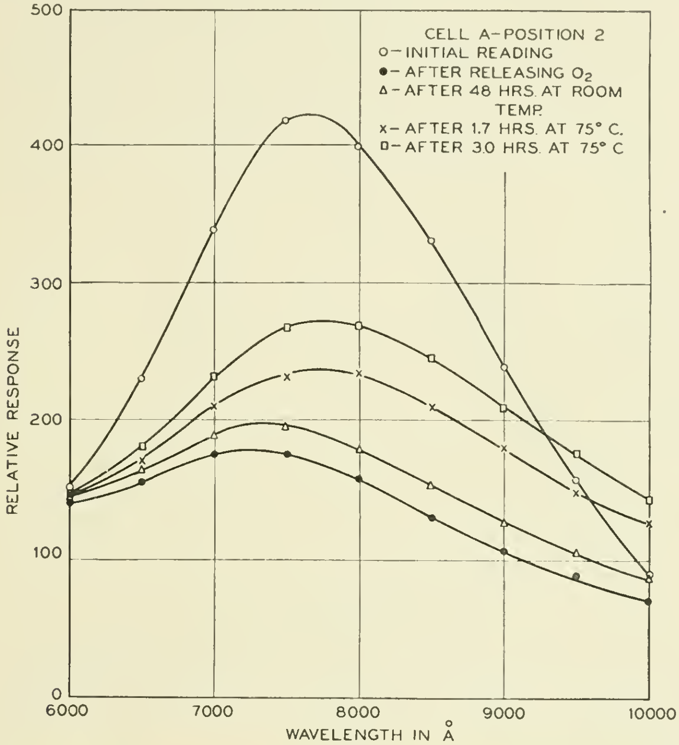


Fig. 9—Relative spectral response to equal energy for cell *A*, position 2.

In cell *B*, Figs. 10 and 11, which we have estimated to be treated with enough oxygen to oxidize a complete atomic layer of cæsium, the selective maximum disappeared at both ends of the cathode and only reappeared as a slight indication in Fig. 11 after prolonged heating at 75° C. On baking for five minutes at 200° C. the selective response reappeared prominently with a maximum at 7,330 Å. but with a loss in response at 9,000 Å. and 10,000 Å. sufficient to involve a decrease in integral sensitivity as shown in Table 2. Further baking involved a further rise in the maximum and a considerable increase in response at 9,000 Å.

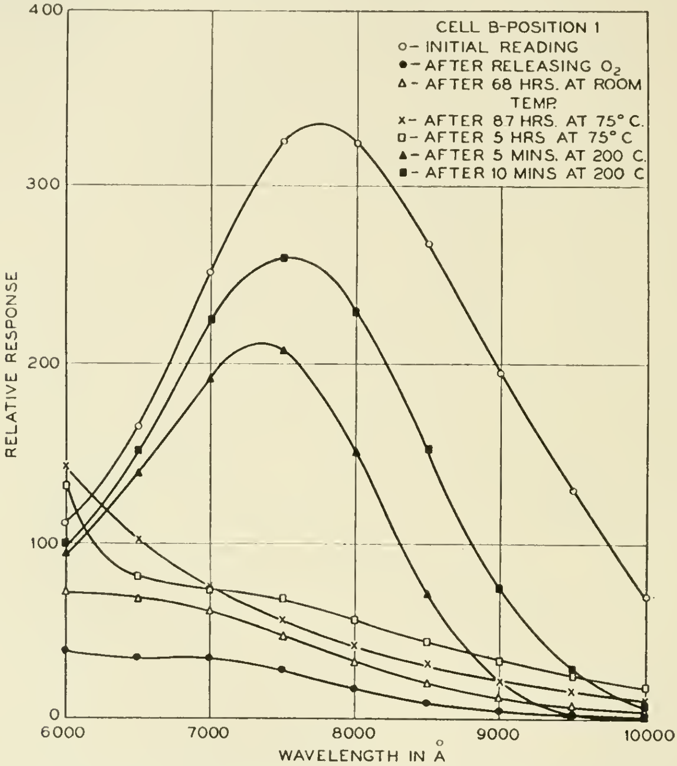


Fig. 10—Relative spectral response to equal energy for cell *B*, position 1.

In cells *C* and *D*, Figs. 12 and 13, which were treated with large amounts of oxygen, the response fell to very low values, rising only slightly on baking at 75° C. But on baking at 200° C. the selective maximum reappeared but at a longer wave-length than in the initial state, and the spectral response at 10,000 Å. was higher than it was at the beginning.

In all cases the recovery is easily interpreted as attendant upon a diffusion of free cæsium from the matrix to the surface film, furnishing a sequence of states with increasing film thickness, and approaching the initial state of each cell, but we can not estimate the quantitative differences in film thickness from state to state on account of the uncertain nature of the oxidation and diffusion processes.

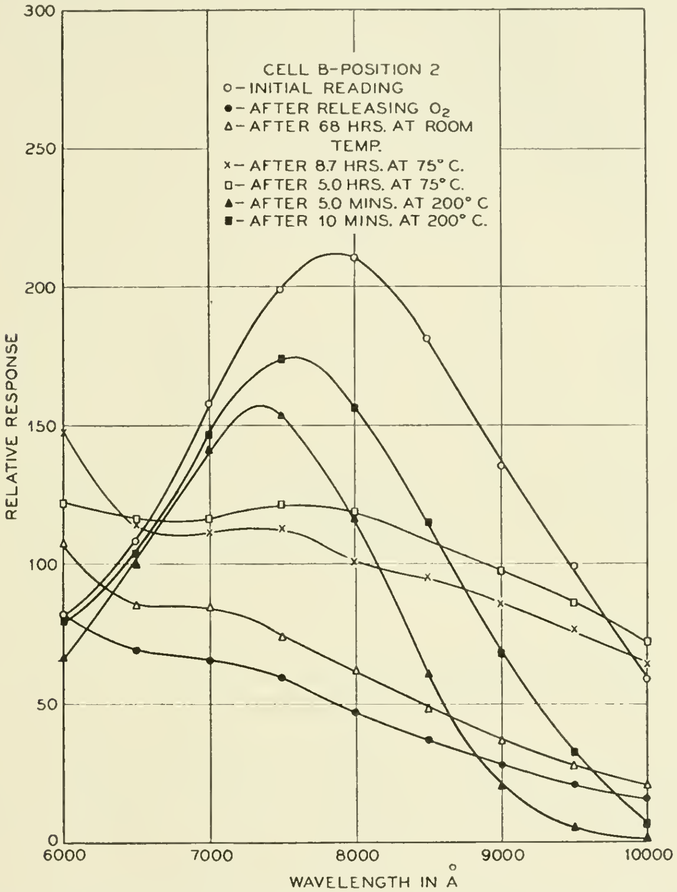


Fig. 11—Relative spectral response to equal energy for cell *B*, position 2.

The extension of this sequence of states to those with film thicknesses greater than that of the initial state was made by the deposition of cæsium upon the cathode surfaces of cells constructed as already shown in Fig. 7. When the cell was shaken and the steel ball made to break the glass bulb containing cæsium, cæsium vapor filled the glass cylinder and diffused slowly through the orifice. Data on three cells are presented in which rates of deposition of free cæsium differing by a factor of 700 were obtained. The sizes of the orifices differed and positions on each cathode at different distances from the orifice were observed. Measurements of spectral response were made with the light incident upon each of the four positions on the cathode shown in Fig. 7.

From simple kinetic considerations we may derive a formula for the rate of deposition upon a surface of the vapor effusing from the orifice:

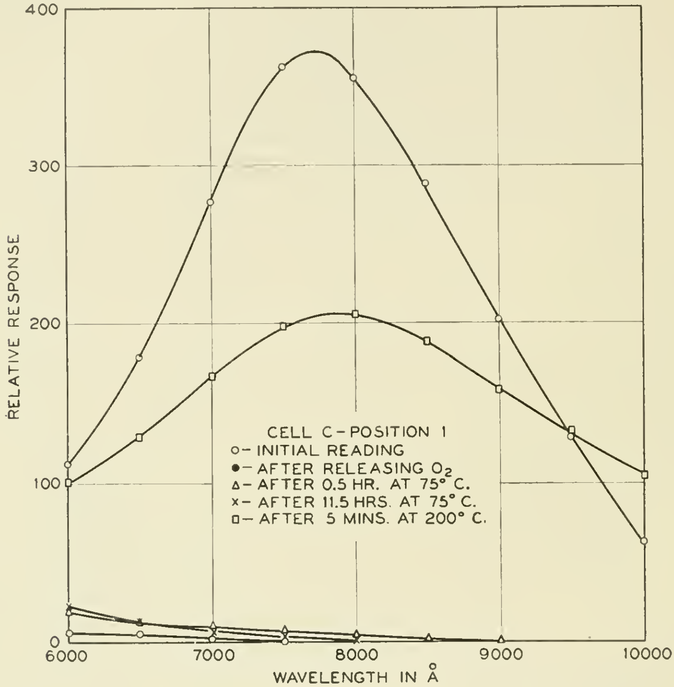


Fig. 12—Relative spectral response to equal energy for cell C, position 1.

$$W = 0.01857\sqrt{M/T}\,\frac{\sigma p_{\text{mm.}}}{r^2}\cos\theta\cos\Psi,$$

where W is the weight deposited per cm.^{2} per second (in grams), M is the molecular weight of the vapor, σ is the area of the orifice, $p_{\text{mm.}}$ is the pressure within the cylinder in millimeters of mercury (i.e. the vapor pressure of cæsium [11]), r is the distance from the orifice to the surface, θ is the angle between r and the normal to the plane

[11] Int. Crit. Tables III, 205.

TABLE 3

Cell Number	σ	t	Position	W_{Cs}	Layers/Hour
E	0.0107cm.^{2}	29.0° C.	2	0.285μgm. cm.$^{-2}$ hr.$^{-1}$	3.15
F	0.000396	28.0	1	0.000407	0.00450
			2	0.0082	0.0906
G	0.00155	31.0	1	0.00218	0.0242
			2	0.044	0.487

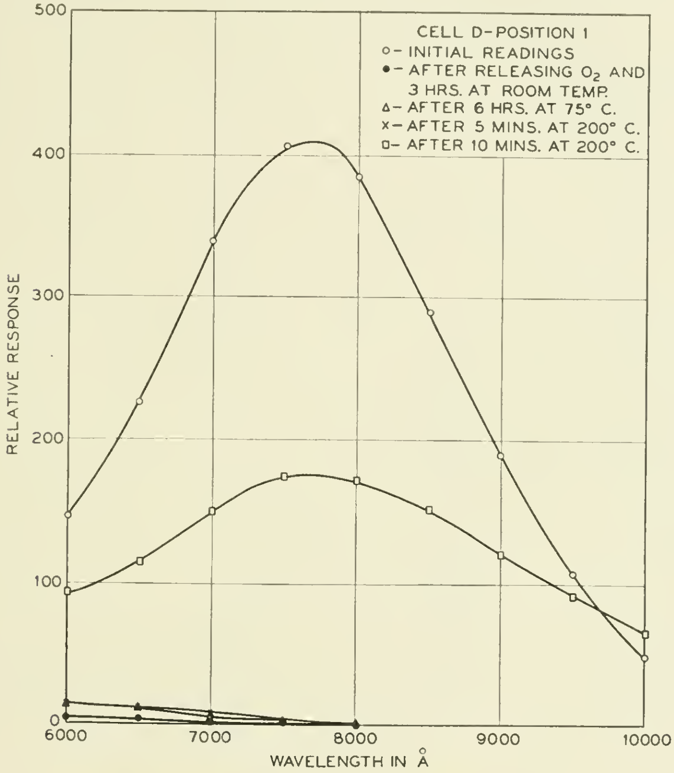


Fig. 13—Relative spectral response to equal energy for cell *D*, position 1.

of the orifice, and Ψ is the angle between r and the normal to the surface. The constants for the three cells presented and the rates of deposition of cæsium at the various positions are shown in Table 3, where t is the mean temperature during the time of observation, and W_{Cs} is the rate of deposition of cæsium in micrograms per square centimeter per hour. In the last column are shown the equivalent number of atomic layers of cæsium that would be deposited per hour upon a plane surface. As before, these values must be divided by approximately four to correct for the increase in surface on roughening, but it seems preferable to present the experimental results in terms of measured rather than estimated quantities, so numbers of layers of cæsium deposited will be presented as upon a smooth surface.

In cell *E*, which had the largest orifice, the changes were so rapid that only one position could be followed. In Fig. 14 the response at each wave-length is plotted against the time on a logarithmic scale.

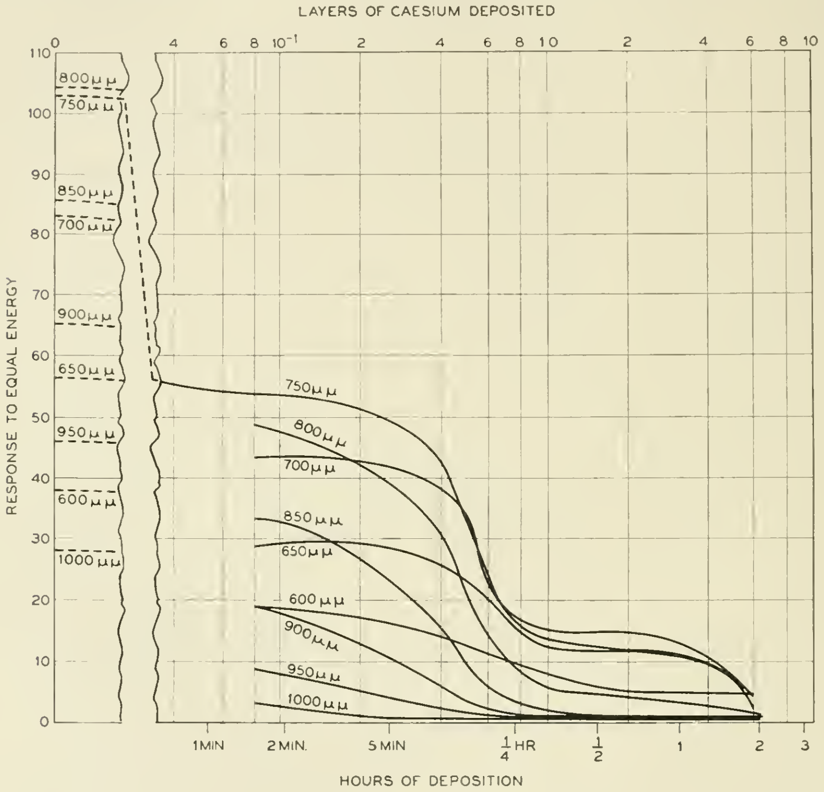


Fig. 14—Spectral response changes in cell *E*.

The response decreased at all wave-lengths, but the effect increased progressively with increasing wave-length. After two hours the wave-length of maximum response had shifted from 7,750 Å. to 6,000 Å.

With slower rates of deposition, initial effects become apparent. In Fig. 15 are shown the results for position 1 of cell *F* which involved the slowest rate of deposition. For wave-lengths shorter than 9,000 Å. there is first a rise in response to a maximum at between 10 and 20 hours of deposition which is most pronounced at 7,500 Å. This is followed by a decay which is again most rapid for the longer wave-lengths. The responses for 9,500 Å. and 10,000 Å. show no rise and decrease steadily after 10 hours. The responses at these wave-lengths were probably at their maxima initially, while the responses at all shorter wave-lengths reach their maxima with increasing time in the order of a progression of wave-lengths from infra-red to visible light. Calculation from the constants in Table 3 shows the maximum response at 7,500 Å. to correspond to the deposition of approximately one tenth of an atomic layer, as may be seen from the scale of abscissæ at the top of the figure.

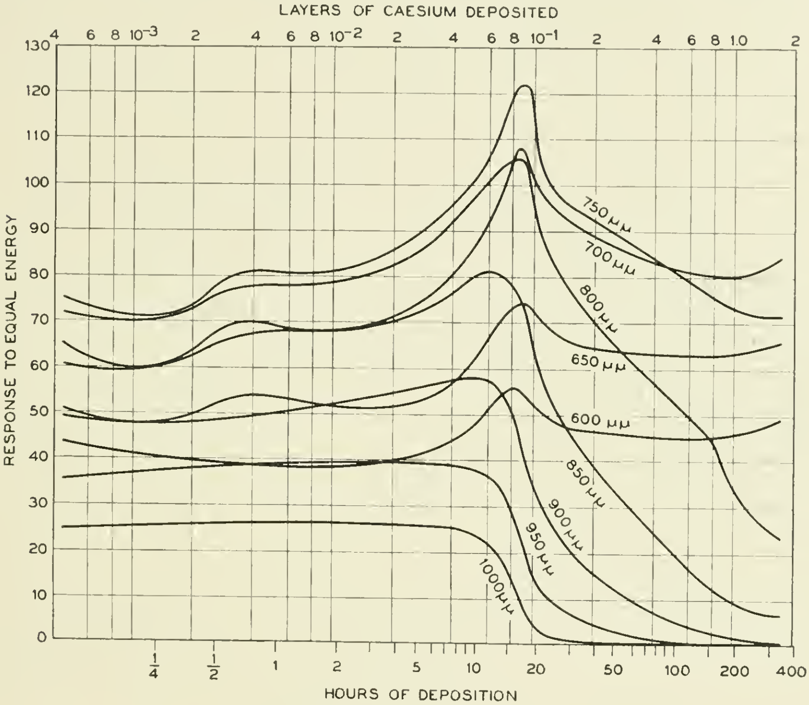


Fig. 15—Spectral response changes in cell *F*, position 1.

For position 2 of this same cell, where the deposition was 20 times as rapid, the curves show a great similarity in every respect including the proportionality of the time scales up to the point corresponding to one tenth of a layer, as may be seen in Fig. 16. Then the maxima spread out and the falling off of response is relatively slower than that found in Fig. 15 as expressed in terms of layers deposited. In fact the two sets of curves become more comparable as expressed in terms of the time scale until after 400 hours the two distributions of spectral response are nearly identical. This final identity may be interpreted in view of the fact that position 2 has received roughly 30 atomic layers of cæsium. For such a thickness the vapor pressure would be expected to approach that of bulk cæsium. At this temperature the rate of vaporization from a bulk cæsium surface would exceed the rate of deposition at this point from the orifice. So cæsium should obviously evaporate from position 2 and condense on the less saturated areas such as position 1. Eventually the thickness of the deposited layer should be equalized over the entire surface.

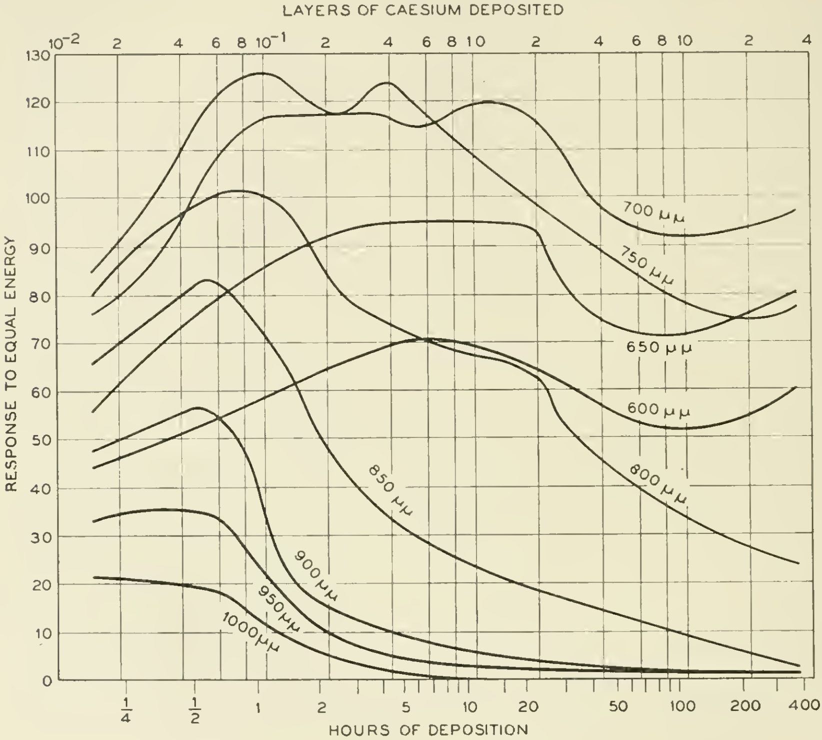


Fig. 16—Spectral response changes in cell *F*, position 2.

The same qualitative description covers the phenomena observed at the other two positions of the same cell which are not shown. The rate of deposition was slow at position 3 and rapid at position 4. The behavior of all four positions fits into a coherent scheme where the initial sections of the four groups of curves are similar with respect to the amount of cæsium deposited and the final sections approach similarity with respect to the elapsed time.

In cell *G* similar qualitative results are apparent. The cell was not as active initially as cell *F*, and the rises in response are not so prominent. For the slow position 1, Fig. 17, the maxima in the response curves are lower and broader than for the slow position of cell *F*, Fig. 15. But the maximum response for a wave-length of 7,500 Å. occurs again after the deposition of approximately one tenth of a layer of cæsium, though in Fig. 17 the maximum response at 8,000 Å. is greater than the maximum response at 7,500 Å. In position 2, Fig. 18, which had the fastest rate of deposition the maxima are either entirely absent or else displaced to a position corresponding to about 0.5 of an atomic layer where there are slight indications of maxima for the wave-lengths 7,500 Å. to 6,000 Å.

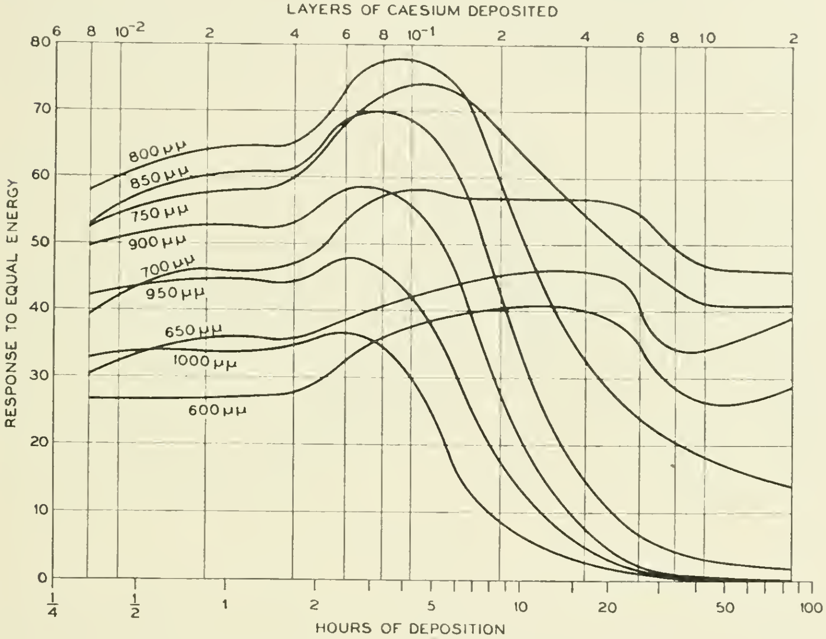


Fig. 17—Spectral response changes in cell *G*, position 1.

If we again refer to Fig. 14, which shows curves for the very rapid deposition of cæsium, it appears in this case that after the deposition of one tenth of a layer the response has decreased markedly for light of all wave-lengths. There appears to be a systematic variation of behavior in that the maxima are very sharp for the slow rates of deposition and are progressively wider and less prominent as the rate of deposition is increased until for very rapid deposition the response for all wave-lengths decreases rapidly from the beginning. In this connection we may recall from the oxidation experiments that a slight amount of diffusion or surface rearrangement may occur in such time intervals as are available in the experiments involving slow deposition of cæsium.

But the behavior of the different cells and even different positions in the same cell may not be closely compared, for their initial states were by no means identical. In each of these cells there was a difference in the colors of the two ends of the cathode due to a distortion of the glow discharge in the quantitative oxidation step of their preparation which was due to the special geometry of these cells

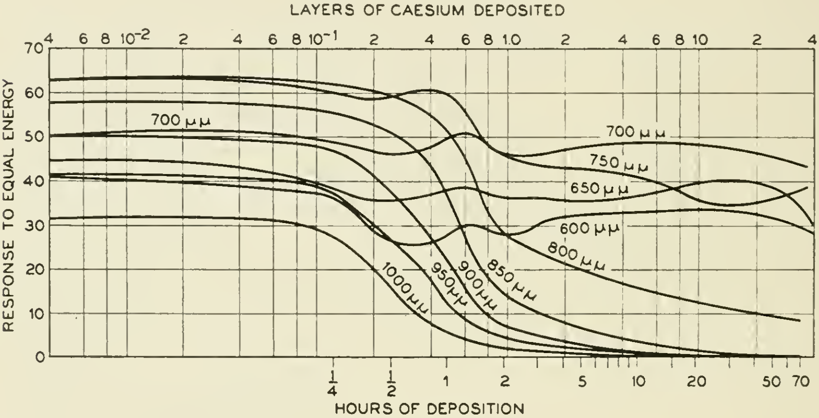


Fig. 18—Spectral response changes in cell *G*, position 2.

But certain gross conclusions may be drawn. In all cases the addition of a few tenths of a layer of free cæsium caused profound modification in the spectral response characteristics. If we regard the results for the slower rates of deposition as more typical of steady state conditions of the cathode surface, it appears that we may increase the response in the region about 7,500 Å. at the expense of the response at and beyond 10,000 Å. by the addition of a few tenths of a layer of free cæsium. It should be remarked again, however, that in view of the roughness of the surface these quantities should be reduced by a factor of four, and the critical amount of cæsium may well be less than even a tenth of a layer on the actual surface. Further addition of cæsium causes a decrease in sensitivity at all wave-lengths, and this effect is increasingly pronounced towards the infra-red end of the spectrum.

In Fig. 19 are shown spectral distribution curves of the response to equal energy with different amounts of deposited cæsium selected from the data shown in Fig. 15. These curves show in more conventional fashion the shift in the position and height of the spectral maximum. This illustrates the general behavior of the equilibrium surface upon the addition of further cæsium. There is an increase in response at the spectral maximum upon the addition of approximately one tenth of a layer of cæsium. As further cæsium is added the height of the spectral maximum progressively decreases. The wave-length of the maximum spectral response and the wave-length of the photoelectric threshold both recede progressively toward the ultraviolet upon the addition of free cæsium to the normal surface.

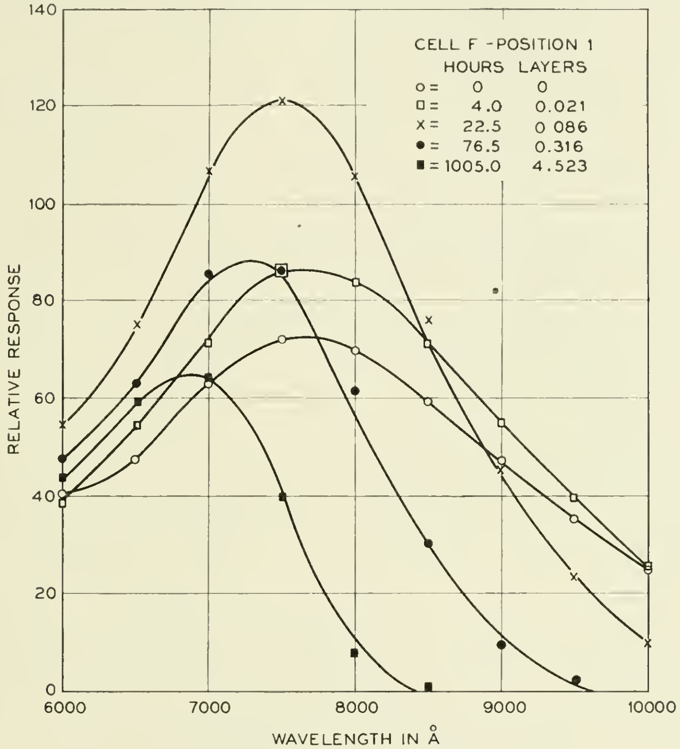


Fig. 19—Relative spectral response to equal energy for cell *F*, position 1.

Combining the results from the experiments where oxygen was admitted to the cells and those where additional cæsium was deposited upon the cathode, we have a sequence of states with increasing thickness of the cæsium thin film upon the cathode surface. If we take curves for the spectral response of various states obtained by oxidation followed by diffusion at room temperature and at 75° C., and then include curves for the spectral response of various states obtained by the slow deposition of cæsium we may obtain a composite family of curves such as are sketched in Fig. 20 to show the whole sequence of response characteristics whose parameter is the thickness of the surface film of free cæsium. In obtaining curves for this composite progression of states we have not considered surfaces after they were baked at 200° C. because of the likelihood at this temperature of surface changes other than in cæsium concentration alone.

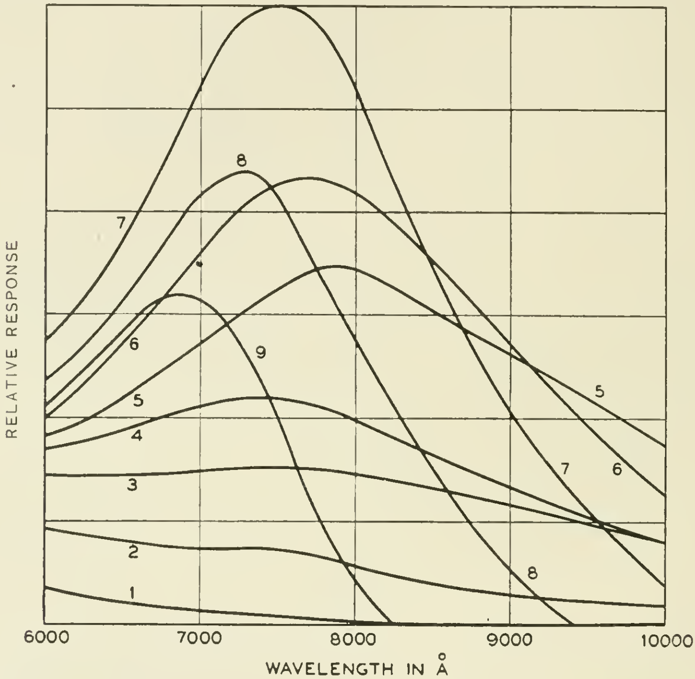


Fig. 20—Relative spectral response to equal energy of a cæsium-oxygen-silver cell as an idealized function of film thickness.

When the amount of cæsium in the surface is far below the normal (as after admitting oxygen equivalent to six layers of cæsium) we have a spectral response which decreases continuously with increasing wave-length, as shown in curve 1 of Fig. 20. This is the same type of response as obtained by Koller[6] from cells prepared by treating with cæsium a silvered bulb with an adsorbed film of oxygen which gave a type of surface designated by him as "Cs–O–Ag." With increasing film thickness the response increases at all wave-lengths

[6] Loc. cit.

and a selective maximum appears as in curves 3 and 4. Then its height rises and its position moves to longer wave-lengths until the maximum occurs at 8,000 Å. as in curve 5, and the response at 10,000 Å. attains its highest value. From a consideration of prevalent concepts of thin-film effects in thermionic and photoelectric phenomena one should expect the surface in this state, which has apparently the maximum value of the photoelectric threshold, to be covered with a monatomic film of free cæsium. As we continue increasing the amount of cæsium in the film, the height of the maximum still rises but its position recedes to 7,500 Å. and the response at 10,000 Å. falls as represented in curves 6 and 7. In the transition from curve 5 to curve 7 the gain in response at 7,500 Å. is offset by the loss in response at and beyond 10,000 Å. As a result the integral sensitivity to tungsten light is sensibly constant and the optimum states for maximum integral sensitivity lie within this range. With further increase in cæsium film thickness the height of the maximum drops, and the positions of both the selective maximum and the photoelectric threshold recede towards shorter wave-lengths as in curves 8 and 9. Eventually we should doubtless approach the spectral response curve for bulk cæsium [12] with the selective maximum at 5,400 Å. and the photoelectric threshold at 7,100 Å.

The large changes in spectral response following the deposition of minute amounts of cæsium upon the cathode surface are further evidence that the photoelectric effect is a phenomenon of the superficial surface only and is conditioned mainly by the thickness of the surface film of free cæsium. Although the cathode surface as a whole is prepared by irreversible chemical reactions which are arrested before their completion, the film thickness of the standard surface is determined by the concentration of free cæsium in the underlying matrix and is maintained by a diffusion equilibrium. In the preparation of the standard surface this partition is first determined at 220° C. in the heat treatment. Stabilization involves a readjustment of the surface to equilibrium with the adjacent portion of the matrix at room temperature. But the attainment of complete equilibrium at room temperature takes so long that many fluctuations of the ambient temperature intervene and the actual history of such a photoelectric surface involves a long series of diffusion waves accompanied by small fluctuations in the thickness of the surface film which determines its photoelectric behavior.

We wish to express our appreciation of the assistance of Mr. M. F. Jameson in these experiments. Acknowledgment is also due to Mr. H. W. Hermance for the chemical analyses, and to Miss A. K. Marshall for the microphotographs.

[12] Miss Seiler, *Astrophys. J.*, **52**, 129 (1920).

Two-Way Radio Telephone Circuits *

By S. B. WRIGHT and D. MITCHELL

This paper deals with the problems of joining long-distance radio telephone transmission paths to the ordinary telephone plant. It gives the possibilities and limitations of various methods of two-way operation of such circuits where the radio channels employ either long or short waves. It also describes the special terminal apparatus for switching the transmission paths under control of voice currents and lists the advantages of using voice-operated devices.

RADIO telephone circuits are now in regular use between New York and London, New York and Buenos Aires, San Francisco and Honolulu, and many other points. At each end of such a circuit there are a transmitting radio station and a receiving radio station, usually geographically separated. The radio provides two one-way transmission paths. The circuit is completed by means of one-way wire lines which connect the radio sending and receiving stations to a common point. At this common point some rather intricate apparatus is called for in order to permit switching of the circuit to the wire telephone plants at the two terminals. This paper explains why this intricate apparatus is necessary even for the comparatively simple case of short-wave radio circuits which use different frequency bands to transmit in opposite directions. It also describes the latest form of this terminal apparatus in which provision is made for certain switching of privacy apparatus by means of which an important saving is made in the amount of privacy apparatus required. The original form of this apparatus is described in an earlier paper.[1]

TRANSMISSION PATHS

Radio telephone circuits may employ the same frequency band for transmission in the two directions or they may employ separate bands. The present long-wave telephone circuit between New York and London is of the first type, while most existing short-wave circuits are of the second.

A short-wave circuit, using separate frequency bands, is shown in its simplest form in Fig. 1. It is formed of two sets of terminal appara-

---

* Presented at I. R. E. Convention, Pittsburgh, Pa., April 7–9, 1932. Published in *Proc. I. R. E.*, July, 1932.

[1] S. B. Wright and H. C. Silent, "New York-London Telephone Circuit," *Bell Sys. Tech. Jour.*, 6, 736–749; October, 1927.

tus connected by two one-way channels, each of which consists of a transmitting wire line, a radio link and a receiving wire line. The function of the terminal or "combining" apparatus is to tie together these two one-way paths in such a manner that they may be connected at the switching centers to various telephone subscribers via the usual telephone circuits.

When the United States subscriber, designated as *A* in Fig. 1, talks, electrical waves set up by his voice pass over a wire line to a toll office. They then divide in a hybrid set. Part of the energy is dissipated in the output of a receiving repeater, and part is amplified by a transmitting repeater and passes over a wire line to a radio transmitter, as indicated in the upper transmission path of Fig. 1.

The waves are then amplified and transformed into radio-frequency energy and radiated. Some energy is picked up by a distant radio receiver, amplified, and transformed back into voice-frequency energy which passes over a wire line to the overseas terminal. The receiving repeater at this point makes up for the loss of the receiving wire line. From its output the waves pass into a hybrid set, part being dissipated in the network and the other part going through the toll-office to the overseas subscriber *B*. Due to the imperfect balance between the subscriber's line and the network, a portion of this energy will be returned over the lower transmission path to the United States subscriber *A* as echo.

The action when the overseas subscriber *B* talks is substantially the same as that described above except that the useful speech waves pass over the lower transmission path.

In long-wave radio circuits the scarcity of suitable radio channels makes it highly desirable to use the same frequency band for transmission in both directions. This results in two additional radio paths becoming important, namely, those between the radio transmitter and the radio receiver at each end of the circuit. By using specially directive antenna arrangements, transmission over these paths may be partly balanced out. In practice, this balance cannot be made very effective in reducing the relative importance of these paths without sacrificing materially the receiving directivity against natural radio noise. The effect of these added transmission paths is to make the transmission problem more difficult, as will be explained.

TRANSMISSION CHARACTERISTICS

Returning to consideration of the simple four-wire set-up involved in short-wave operation, the transmission characteristics of the circuit evidently depend on the sum of the effects of the radio and wire line

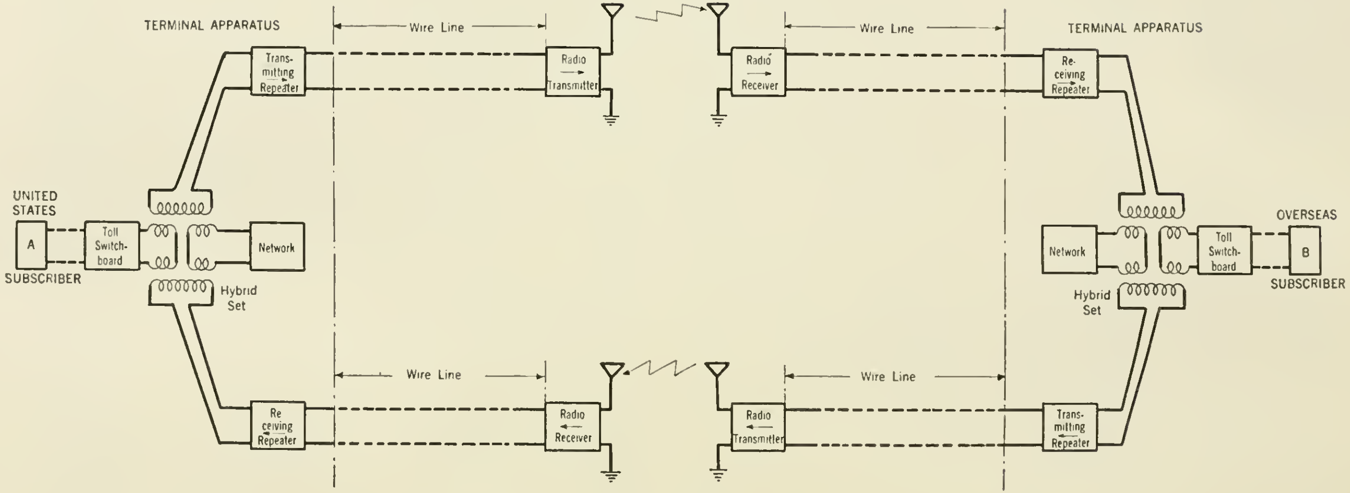


Fig. 1—Two separate radio links connected as a four-wire circuit.

portions. In view of the relatively higher cost and greater length of the radio links, the highest grade wire circuits available are usually justified so that, in general, they should not be allowed to contribute much transmission impairment. In general, the added delay introduced by these wire lines is their most important effect.

If the radio links are quite stable and fairly free from atmospheric disturbance, the circuit may be operated like four-wire land telephone circuits. That is, the total amplification in the circuit may be kept at such a value that it never exceeds the total attenuation, and over-all singing will not occur. Four-wire terminating sets or hybrid sets placed at the ends of the circuit, as shown in Fig. 1, are adequate to prevent singing and minimize echo effects in four-wire circuits which have over-all transmission times less than about 0.02 second, provided the net loss from switchboard to switchboard does not become lower than about 5 db.

If the radio links or wire lines are long, the circuit will produce annoying echoes exactly as a four-wire cable circuit will, due to delay or instability, or both. Also, as in the case of four-wire cable circuits, echo effects may be reduced appreciably by voice-operated echo suppressors which block the path of the delayed echoes while the other path is transmitting speech. The possibilities and limitations of this type of device are discussed elsewhere.[2]

When the radio channels are more noisy and/or less stable, the transmission may be greatly improved through more efficient use of the radio links. The noise may be minimized by bringing the speech waves of all talkers, strong or weak, to the same "electrical volume" or strength at the input to the radio transmitter. Thus, practically full modulation may be maintained on the radio transmitter at all times and the ratio of the desired signal to the radio noise kept a maximum. Large changes in gain between the two-wire line and the radio transmitter are necessary to accomplish this result. These changes are made by technical operators who make the adjustments with variable loss devices. An indication of the volume is obtained through the aid of electrical meters called "volume indicators." In practice, the over-all transmission of a long radio circuit may be varied by the technical operators from a 30-db loss to a 30-db gain within a few minutes.

In short-wave circuits, the phenomenon known as "fading" introduces an effect which is of great importance in two-way operation. Where fading results in variations of the entire transmitted band of frequencies, automatic gain control at the radio receiver is effective in maintaining the received volume at a substantially constant value. The gain control is operated by the incoming carrier. When the fading is of the type in which the different frequencies in the transmitted band do not fade simultaneously, the automatic gain control is not so effective and considerable variations in volume out of the receiver may occur in a short time.

---

[2] A. B. Clark and R. C. Mathes, "Echo suppressors for long telephone circuits," *Jour. A. I. E. E.*, **44**, 618–626; June, 1925.

In the long-wave circuit, the variations are too slow to be classed as fading, and occasional manual adjustments of receiver gain result in keeping the volume at the receiving end within about ±5 db.

Because of the gain adjustments to reduce noise, combined with changes in radio receiver gain to compensate for fading or for variations in radio attenuation, "singing" would occur if the hybrid coils and echo suppressor were not augmented by additional means of singing prevention. One way of preventing singing would be to reduce gain in the receiving leg whenever gain was introduced in the transmitting leg of the circuit. Volume penalties to the listener as great as 25 db would frequently be encountered if this were done, and, in addition, considerable agility would be required on the part of the technical operators to keep the circuit adjusted. However, this method would not compensate for gain changes in the radio receivers, so that singing might still occur under unfavorable conditions.

Also, in the case of a long-wave transatlantic circuit, singing could occur over transmission paths between the local radio transmitter and receiver. The volume received from the local transmitter may occasionally be as much as 40 db stronger than that from the distant station if the transmitter and receiver are about 90 miles apart, even though antenna directivity were used at both the transmitters and the receivers. In general, if the receiver gain is adjusted to give the proper volume on the distant station, the amplification in the local radio path is entirely out of reason.

It is therefore necessary to provide other means of preventing singing to maintain optimum transmission conditions.

Vodas

There has been developed for meeting these difficulties an anti-singing voice-operated device known as a "vodas." [3] Fig. 6 shows a radio telephone circuit arranged with a vodas in its simplest form at each end of the circuit. The vodas consists of a transmitting delay circuit, detector, and certain relays, and a receiving delay circuit, detector, and relay. These devices are operated by the voice currents in the circuit so as to keep all singing paths blocked at all times.

[3] Taken from initials of the words "Voice-Operated Device Anti-Singing."

The vodases in Fig. 6 are shown for the condition when no speech is being transmitted. Relay 1 keeps the transmitting circuit blocked so that singing cannot occur around the complete circuit or through a local radio path and terminal. Transmission is free to pass the contacts of relay 2. When the United States subscriber speaks, voice currents go into the transmitting detector and delay circuit. While they are traversing the delay circuit, relays 1 and 2 become operated provided relay 3 has not been operated previously. The operation of relay 1 permits the voice currents to travel on to the radio transmitter. The operation of relay 2 blocks the receiving path and prevents echoes and singing that might otherwise occur when relay 1 is operated.

Upon being received at the distant end, the voice currents operate relay 3 from the receiving detector, thus protecting the transmitting detector and relays against operation by echoes of received speech currents. These echoes are returned from unbalances in the two-wire portion of the connection beyond the terminal. The receiving delay circuit delays the speech long enough to insure complete operation of relay 3 before the echoes return. When the subscriber stops speaking, the relays return to normal.

By adding two more relays to the transmitting side of the vodas, it is possible to save part of the apparatus which is used to increase privacy on the circuit. This saving is made by using the same privacy device for both transmitting and receiving. This is possible provided the action of the privacy device is the same for distorting the voice waves at the transmitting end as it is for restoring them at the receiving end of the circuit. An arrangement of the vodas having this feature, which is now in general use, is illustrated in Fig. 2. The apparatus additional to the simple vodas consists of relays 4 and 5, the privacy device, a hybrid set, and two one-way repeaters. In Fig. 2 this apparatus is labeled "Privacy Switching Circuit."

The action of the device shown in Fig. 2 on transmitting speech waves is as follows: Useful waves from subscriber *A* are impressed on a potentiometer ahead of the transmitting repeater, which is kept adjusted by the technical operator to maintain constant volume at the output of the transmitting repeater. The waves then pass into the vodas where they first reach the transmitting delay circuit and are stored for a short interval. A small part of these waves enters the transmitting detector and operates relays 1, 2, 4, and 5, provided relay 3 has not been operated by the receiving detector previously. The interval of the transmitting delay circuit is several times as great as the operating time of relays 1, 2, 4, and 5 so that initial weak parts of speech syllables may be stored in this delay circuit until stronger parts arriving later have had a chance to operate the relays.

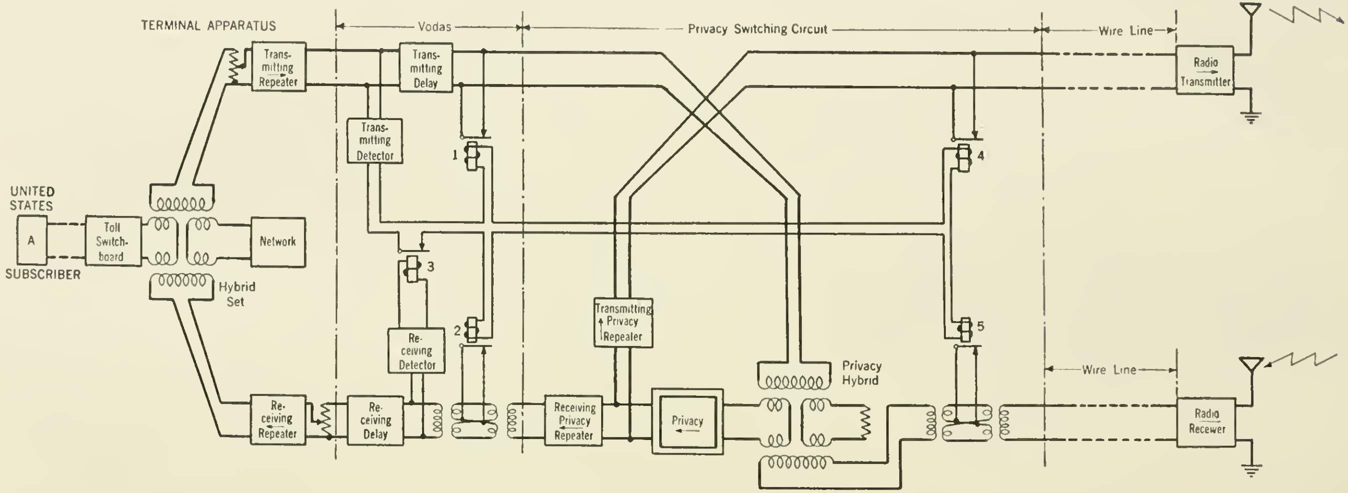


Fig. 2—Voice-operated device used on New York-Buenos Aires and other radiotelephone circuits.

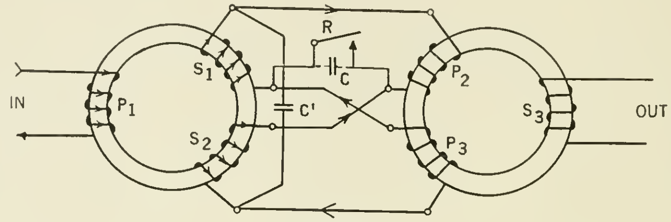
Operation of relay 1 clears a path for the speech waves after they pass through the transmitting delay circuit. From there they pass into the privacy switching circuit traveling downward to the privacy hybrid set. Here they divide and the useful portion travels through the privacy device. At its output the waves divide, part being amplified by the transmitting and part by the receiving privacy repeater. The subscriber at *A* is prevented from hearing echo by the operation of relay 2 which disables the receiving transmission path in a way to be described later. The portion of the waves which travels upward through the transmitting privacy repeater, however, is now free to pass on to the wire line and radio link due to the operation of relay 4. Operation of relay 5 prevents any echo due to direct transmission from the local radio transmitter to the radio receiver from being passed into the privacy device and thus causing distortion of the outgoing waves.

Action of this device on incoming speech waves is as follows: The waves after being detected by the radio receiver travel over the wire line and, provided the transmitting relays have not been previously operated, pass through the first repeating coil combination and into the lower side of the privacy hybrid. There they divide and the useful part passes through the privacy device where it is restored to an intelligible form. At the output of the privacy device, the waves are amplified by both transmitting and receiving privacy repeaters. Any retransmission of these waves from the local transmitter is prevented by relay 4 which is now released. At the output of the receiving privacy repeater the waves pass through the second repeating coil combination and thence to the receiving delay circuit. A part of these waves, if they are strong enough, may operate the receiving detector and thus, relay 3.

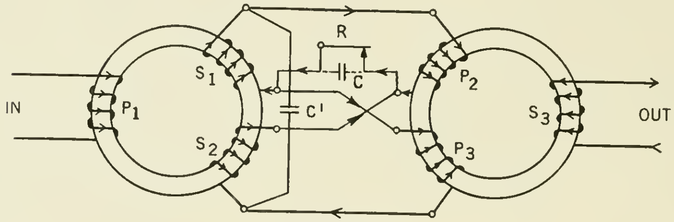
After passing through the receiving delay circuit the speech waves travel on through a receiving potentiometer, receiving repeater, hybrid set, and to the subscriber at *A*.

A small part of these waves may be reflected due to the difference in impedance between the subscriber's line and the network and return over the transmitting branch of the circuit. However, if these waves are strong enough to operate relay 3, they are prevented from operating the transmitting relays. If they are too weak to operate relay 3, an adjustment of receiving volume is made so that they will be too weak to operate the transmitting relays. This adjustment is accomplished by the potentiometer ahead of the receiving repeater. The waves are prevented from passing through the privacy device by relay 1 which is released.

In practice, it is necessary to protect the vodas against operation from noise. This is done by frequency discrimination as well as by amplitude discrimination and the use of artificial delay. The detectors have their input circuits arranged to keep out frequencies that are not essential for speech operation. In addition, their sensitivity is made adjustable. The transmitting detector is generally worked at a value which results in no perceptible loss of intelligibility due to failure of the transmitting relays to operate, at the same time allowing a maximum amount of line noise to be applied without operating the relays falsely. The receiving detector is adjusted frequently by the technical operator so as to obtain the best operation on incoming speech without false operation from radio noise. When the incoming noise is low, relay 3 may be made very sensitive. Any incoming speech which does not operate relay 3 is thus weak, and the receiving volume may be kept high without danger of echoes operating the transmitting relays. When the noise is high, relay 3 is made insensitive, requiring more loss in the echo path and, consequently, lower volume to the listener.



A. SUPPRESSED



B. UNSUPPRESSED

Fig. 3—Repeating coil arrangement for suppressing echoes.

The method of suppressing transmission by opening a single relay contact is illustrated in detail by Fig. 3. In A of this figure, the relay (*R* on the figure) is assumed to be operated so that transmission is suppressed. The voltages induced in windings S_1 and S_2 of the first coil are opposed to each other in a circuit including P_2 and P_3 of the second coil, the resulting flux in the core of this coil being very small. Losses

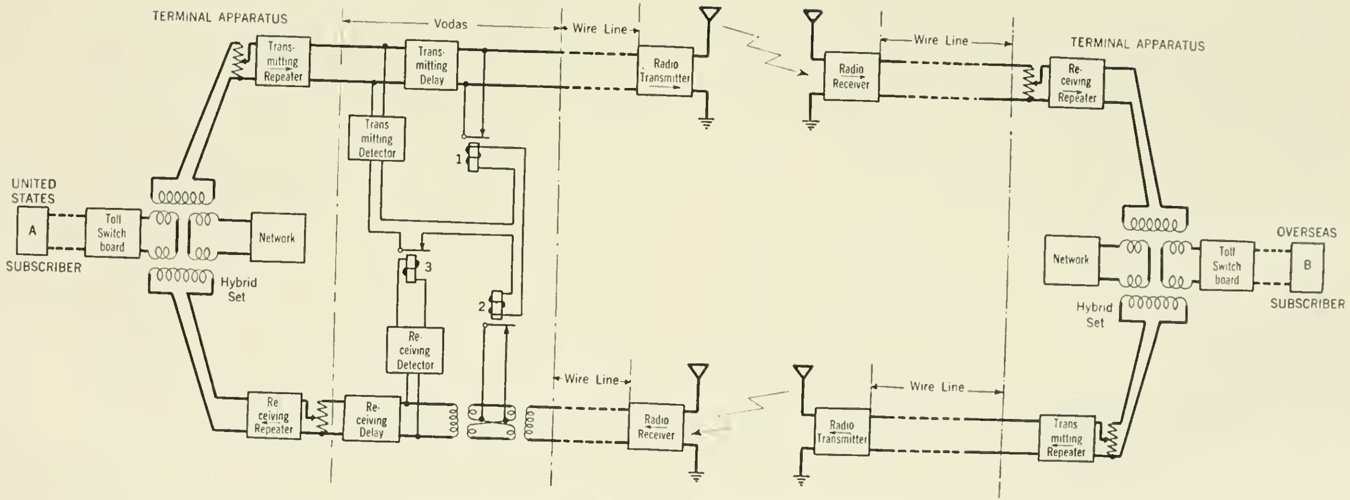


Fig. 4—Radio telephone circuit with vodas at one end only.

as high as 75 db are produced by this arrangement by proper design of coils and adding a small condenser (C') to balance the capacity (C) of the leads to the relay contacts. In practice, a pair of wires is cut off to give the right capacity, and then laced into the cable form. In B of the figure, the circuit is shown in the normal condition. Transmission through the coils is affected only by their normal transition loss. While the windings S_1 and S_2 in A are effectively in series opposing, in B they are connected by separate circuits to the corresponding windings P_2 and P_3, due to the extra path through the relay contact.

OPERATION OF A RADIO TELEPHONE CIRCUIT

Having in the vodas a means for suppressing echoes and singing under extreme conditions (with the additional advantages of suppressing intermediate "cross-transmission" paths), it is important to consider the broader application of such a device to radio telephone circuits. Three cases of operation with anti-singing devices are of interest:

1. Vodas at one end, plain hybrid set at the other.
2. Vodas at one end, echo suppressor (without anti-singing relay) at the other.
3. Vodases at both ends.

1. Vodas at One End, Plain Hybrid Set at the Other

This arrangement is shown in Fig. 4. In this and the next figures, the privacy switching circuit has been omitted for simplicity. The vodas prevents singing and echo effects from unbalances at the *A* end and also prevents the *A* subscriber from hearing echoes.

The disadvantages of these arrangements may be understood by considering the transmission received at the *B* end which has no voice-operated relays. Speech received over the circuit would be returned to the local radio transmitter as an echo or echoes. If a weak talker were connected at the *B* end, the volume control device would amplify these echoes to an appreciable extent. In addition to overloading the radio transmitter, such echo would permit both sides of the conversations to be broadcast from the same station, thus reducing privacy. Radio noise might also be received at *B* and transmitted as echoes to the *A* end of the circuit. In addition, line noise from a two-wire circuit at the *B* end would be freely transmitted to the *A* end, causing a limitation of the sensitivity of relay 3 and consequently a reduction of volume at the *A* end.

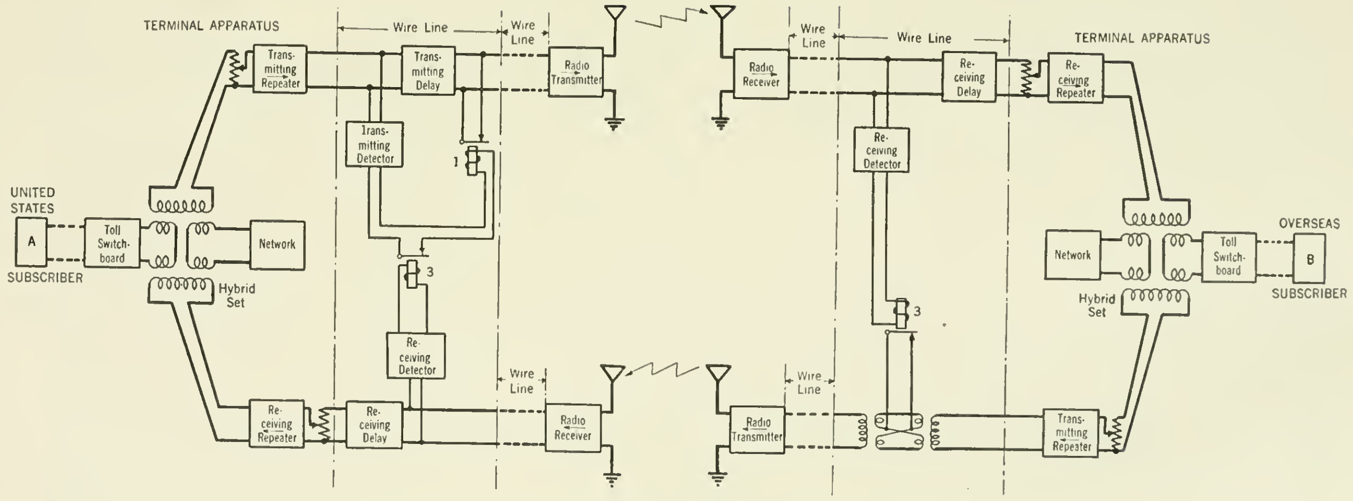


Fig. 5—Radio telephone circuit with vodas at one end and echo suppressor at opposite end.

2. Vodas at One End, Echo Suppressor at the Other

The above possibilities of transmitter overloading and speech reradiation due to echoes might be prevented by adding a voice-operated echo suppressor at the *B* end as shown in Fig. 5. This device would be operated by received speech so as to disable the transmitting branch of the circuit. Its sensitivity would be limited as would also the received volume at one or both ends. It should be noted that if no cross transmission paths existed, relay 2 of the vodas at the *A* end could be omitted when this device is used.

The echo suppressor would not suppress echoes of radio noise or direct transmission of line noise from the *B* end to the receiving relay 3 at the *A* end. Relay 3 would therefore need to have its sensitivity reduced so as not to be operated by these noises. This gives rise to an additional limitation of received volume at the *A* end. The amount of the penalty would, of course, depend on the noise conditions, the talker volume and the echoes in two-wire circuits at the *B* end. Under extreme conditions, the necessary reduction in receiving volume at the *A* end might be as much as 25 db. This is considered to be an important disadvantage inasmuch as conditions at the two ends are not independent and lack of an anti-singing device at one end penalizes the received volume at the other.

Another solution would be to limit the transmitting gain at the *B* end so that the noise transmitted past the echo suppressor would never limit the sensitivity of relay 3 at the *A* end. This would mean that the received signal-to-noise ratio at the *A* end would be reduced, particularly if the taker at the *B* end were weak.

3. Vodases at Both Ends

To summarize, it may be said that anything short of anti-singing devices at both ends does not make the two ends of the circuit independent and may penalize the transmission at the vodas end when there is radio noise or line noise at the end without an anti-singing device. The preferred arrangement is shown in Fig. 6.

RESULTS OF VODAS OPERATION

In general, the results of operation with the vodas have been good and, when radio conditions are favorable, the circuits are not appreciably different from land circuits of comparable length.

Occasionally, the vodases introduce minor difficulties. False operation by noise and simultaneous talking by the two subscribers both tend to cause speech mutilation. The large transmission advantages

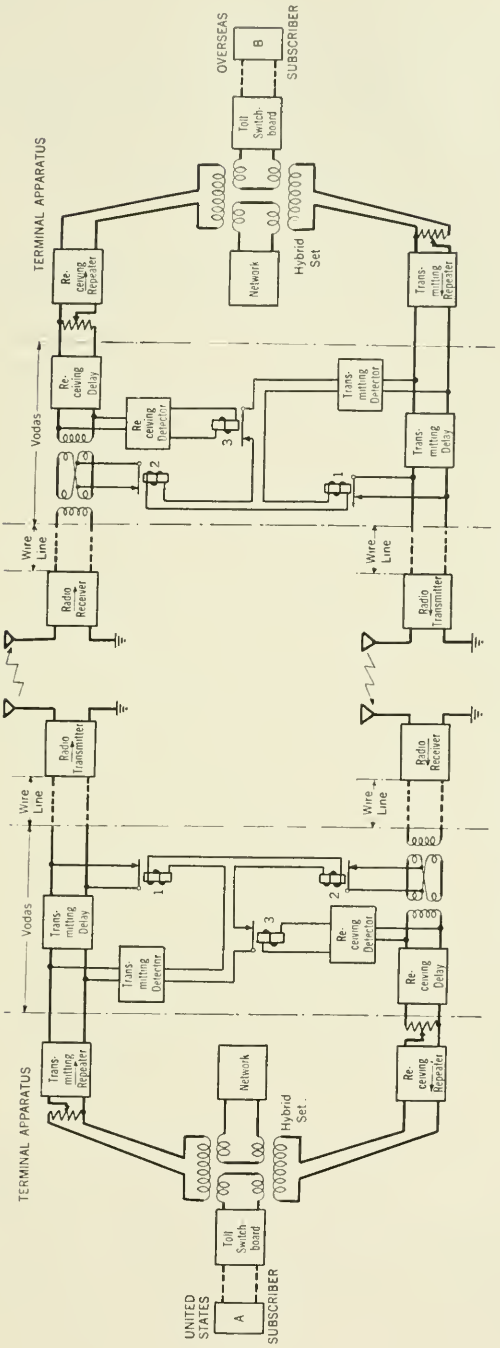


Fig. 6—Radio telephone circuit with vodases at both ends.

afforded by their use greatly outweigh any such troubles. These advantages may be summarized as follows:

1. Suppress echoes and singing which would otherwise be heard due to adjustments to reduce radio noise, instability and cross-transmission paths.
2. Prevent retransmission of echoes which would cause overloading and two-way broadcasting at the radio transmitters.
3. Save privacy apparatus.
4. Permit the telephone listeners to hear louder speech waves.
5. Afford independent technical operation of the two ends of the circuit.

Magnet Steels and Permanent Magnets—Relationships Among Their Magnetic Properties*

By K. L. SCOTT

Introduction

A GENERAL study of the magnetic properties of magnet steels and permanent magnets embraces a number of related problems. Of chief interest to the designer of magnets is the problem of determining the relationship between the open-circuit remanence of a permanent magnet and the various factors which determine its value. These factors include the magnetic characteristics or properties of the steel, as displayed by its hysteresis loop, and the shape and dimensions of the magnet. Of interest to the manufacturer of magnet steel are the questions of how chemical composition, melting practise, and rolling practise affect the magnetic and mechanical properties of the steel. The manufacturer of magnets is interested in the effect of the necessary manufacturing operations, such as heating for hot forming and annealing to increase machinability, upon the flux obtainable in the finished magnet; on the tendency of the steel to warp and crack upon quenching; and on the proper hardening treatment to use.

These interests overlap, of course, and include many special subjects, among which may be listed the manner in which various influences may affect the state of magnetization of a magnet, the correlation of the microstructure and mechanical properties of a magnet steel with its magnetic properties, the determination of a suitable criterion of magnetic quality, the development of accurate and convenient testing equipment, and various other matters of greater or less importance.

It is the purpose of this paper to present data relating to some of the above topics, which have been collected by the writer during the course of several years connection with the manufacture of permanent magnets, both in the laboratory and in the shop.

Symbols and Notations

The following symbols and notations will be used. See Fig. 1. All values of B are intrinsic or ferric induction.

B_{max} = The value of magnetic induction corresponding to the tip of a given hysteresis loop, in gauss.

---

* Presented before A. I. E. E. in New York City, Jan. 26, 1932. Published in abridged form in *Elec. Engg.*, May, 1932.

H_{max} = The value of H corresponding to B_{max}, in oersteds.

B_s = The saturation value of flux density corresponding to an infinite magnetizing force.

B_r = Residual induction, the magnetic induction in a ring or infinitely long straight bar after the value of H has been reduced from H_{max} to O. The value of B at the intersection of the hysteresis loop with the B axis, in gauss.

H_c = Coercive force, the value of H required to reduce B from B_r to O in a ring or infinitely long bar. The value of H at the intersection of the hysteresis loop with the H axis, in oersteds.

B_{rem} = Remanence, the magnetic induction at the magnetic equator of a permanent magnet with no external magnetizing or demagnetizing force. Values of B_{rem} given in this paper are without pole pieces on the magnets, in gauss.

L = The actual developed length of a magnet.

A = The area of cross section of a magnet.

D = The equivalent diameter of a magnet
$= 2\sqrt{A}/\sqrt{\pi}$

L/D = The dimension ratio of a permanent magnet.

$(BH)_{max}$ = The maximum value of the product of the coordinates of the demagnetization curve for a given steel. The demagnetization curve is the portion of the hysteresis loop between B_r and H_c.

T = Hardening temperature, absolute scale.

T_0 = Optimum hardening temperature, absolute scale.

Factors that Determine the Remanence of a Permanent Magnet

Of the subjects enumerated in the introduction, the first is the one that has been of greatest technical interest to the writer. During the course of several laboratory investigations connected with shop problems, advantage was taken of the opportunity to secure data which might be used for the purpose of making generalizations regarding relations among magnetic properties of magnet steels and permanent magnets.

The specific problem was to find out exactly the way in which the remanence of a magnet is related to its shape and dimensions as well as to the magnetic properties of the material of which it is made. Although it requires the complete family of hysteresis loops for a given piece of magnet steel to portray all of its various magnetic characteristics, it was felt that possibly the combined influence upon the remanence of a magnet of the sum of these characteristics could be expressed in terms of the values of B_r and H_c for the material. Such an assumption, if valid, would simplify the problem considerably.

It is true that in the literature on the subject there are several papers [1, 5, 6] describing more or less accurate methods by means of which a designer may predict the flux of a given magnet provided the demagnetization curve for the steel is known. But these methods are laborious, and unless a complete redesign is worked out for each case, they do not answer such questions, for example, as how much a certain change in B_r or H_c will affect the remanence of a given magnet, other factors remaining constant.

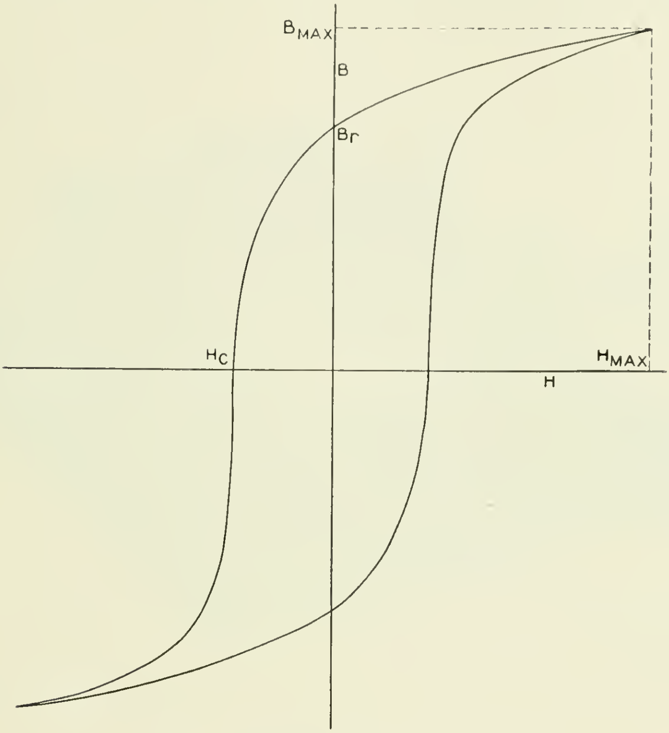


Fig. 1—Hysteresis loop of magnet steel.

Because the flux value of a magnet is affected by shape as well as other factors, and because magnets are made in a bewildering variety of shapes, it was decided to confine the investigations to straight bar magnets and leave the matter of the effect of shape on flux for a later study.

---

[1] For references see end of paper.

Test Specimens. Accordingly, a large number of straight bars of magnet steel were used for test specimens. Each of the bars was 12 inches in length. In the lot were bars of ⅛ by 1½ in., ¼ by ¼ in., 3/16 by ¾ in., ¼ by 1¼ in., ¾ by ⅝ in., and ⅜ by ½ in. cross sections. All of the types of steel available were used, including 0.9 to 3.5 per cent chromium steel, 0.85 per cent manganese steel, 5 per cent tungsten steel, and 20, 25 and 36 per cent cobalt steels. Along with this range in composition a range in coercive force of from 40 oersteds to 260 oersteds was available, accompanied by a variation in residual induction of from 6,000 gauss to 11,000 gauss.

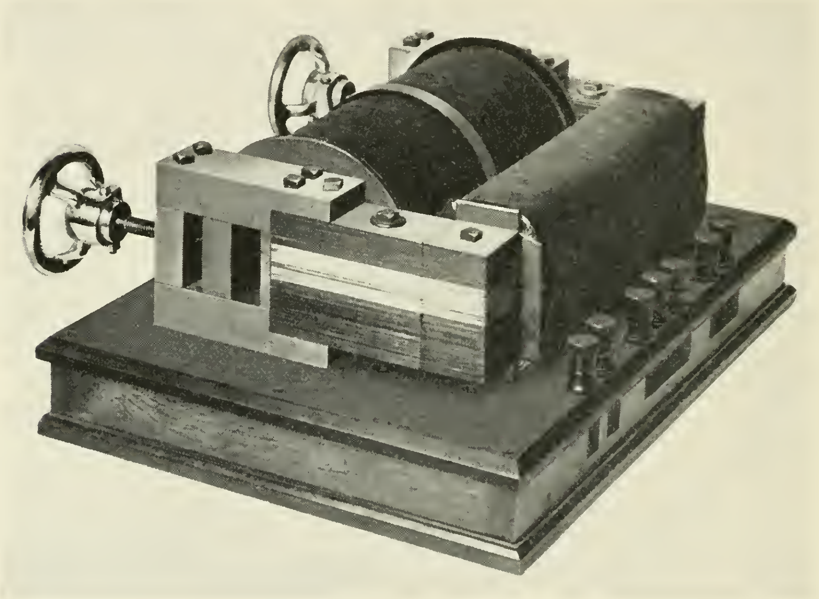


Fig. 2—The Babbitt permeameter. Used for making magnetic measurements on any kind of magnet steel.

Experimental Procedure. The test bars were hardened in the manner appropriate to each type of steel, except that some bars were purposely quenched from a high temperature in order to secure low values of B_r. After hardening, the demagnetization curve for each bar was determined by means of permeameter measurements, using the Babbitt[7] permeameter and a Grassot fluxmeter with lamp and scale. The control circuit of this permeameter is the conventional circuit for a ring test, as shown in Fig. 3. The magnetic data are all comparable since accurate measurements may be made with this permeameter on all of the above mentioned types of magnet steel.

Following the permeameter measurements, each bar was magnetized as a straight bar magnet in an air core solenoid. Field strengths of over 1,000 oersteds were applied to each bar. Upon removal from the solenoid the flux density at the middle of the bar was measured by means of a search coil and the Grassot fluxmeter with lamp and scale. It is the flux density measured in this way that is referred to in this discussion as the remanence of a magnet, or B_{rem}.

After the measurements at 12-in. lengths were completed, each bar was cut in succession to 10, 8, 6, 4, and 2-in. lengths by removing 1 inch of material from each end except that in the case of some of the bars of larger cross-section the lengths of 10, 6, and 2 inches were omitted.

Remanence measurements were made at each of the above lengths.

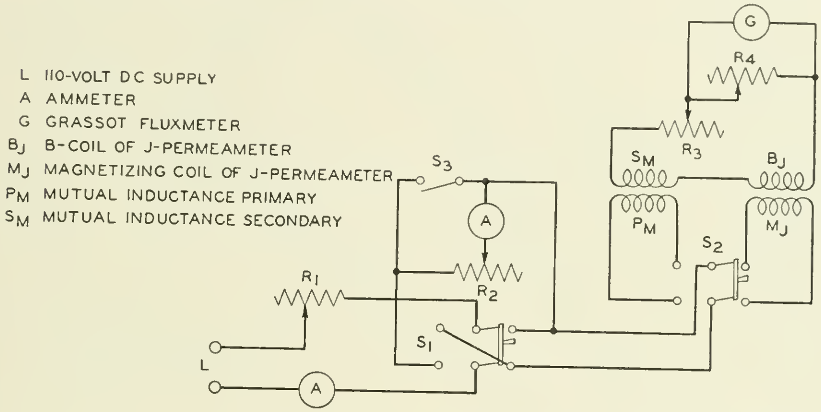

L 110-VOLT DC SUPPLY
A AMMETER
G GRASSOT FLUXMETER
B_J B-COIL OF J-PERMEAMETER
M_J MAGNETIZING COIL OF J-PERMEAMETER
P_M MUTUAL INDUCTANCE PRIMARY
S_M MUTUAL INDUCTANCE SECONDARY
G
R_4
R_3
S_M
B_J
S_3
A
P_M
M_J
S_2
R_1
R_2
S_1
L
A

Fig. 3—Control circuit of the Babbitt permeameter. Note that the connections are those of a simple ring test.

The bars were remagnetized for each new measurement of B_{rem}.

The foregoing experiments made available for analysis a wide range of values of the associated variables B_{rem}, B_r, H_c, length L, and cross-sectional area A.

Analysis of Data. The values of remanence *versus* length for a number of the straight bar magnets are plotted in Fig. 4 to show the range of values existing in different bar magnets of the same length, when fully magnetized. The differences in B_{rem} for a particular length of bar are due to the differences in B_r, H_c, and cross-section of the various samples.

The chief characteristic that the curves of Fig. 4 have in common is a resemblance in shape to a normal induction curve and an asymptotic approach to limiting values of B_{rem} as the lengths increase.

In Fig. 5 is shown the result of plotting B_{rem}/B_r *vs.* $(L/D)\sqrt{H_c}(T/T_0)^2$ for the same set of bars. This combination of variables was arrived at purely by a cut and try method, but as shown, a surprisingly good correlation is obtained.

It should be mentioned in connection with Fig. 4 that it was found that the data relating to some of the bars did not fit the curve satisfactorily if B_{rem}/B_r was plotted against $L\sqrt{H_c}/D$, but in each case where this was so the bar had been quenched from above or below the optimum hardening temperature. It was found, however, that the data from these bars could be correlated with those from the properly hardened bars if B_{rem}/B_r were plotted against $(L/D)\sqrt{H_c}(T/T_0)^2$ in which T is the actual, and T_0 is the optimum hardening temperature on the absolute scale. The ratio T/T_0 is, of course, equal to unity for bars properly hardened. The definition of the optimum hardening temperature will be given later.

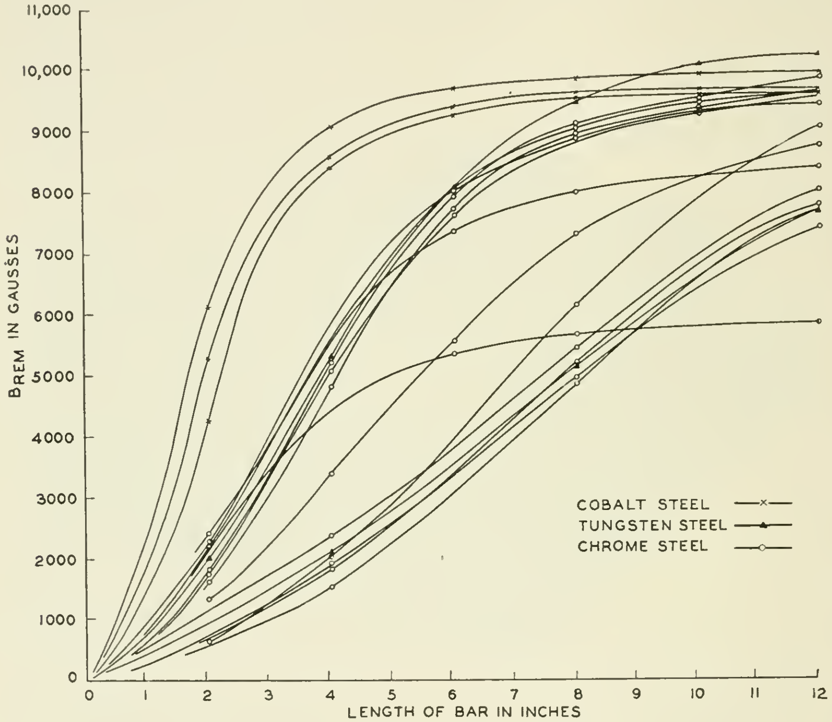


Fig. 4—Curves of remanence *vs.* length for straight bar magnets of different cross-sections and materials.

After this empirical correlation was obtained, it was suggested that B_{rem}/B_r *vs.* $L\sqrt{H_c}/D\sqrt{B_r}$ would be preferable as a choice of variables, from theoretical considerations based upon the assumption of uniform magnetization in the magnets. Although it can be shown that this assumption is not fulfilled in the actual case of straight bar magnets, this method of plotting gives as good a correlation as is shown in Fig. 5. Such a result might be expected, since the values of B_r commonly encountered range from 9,000 gauss to 11,000 gauss, thus differing from 10,000 gauss by 10 per cent at most. The square root of B_r, therefore, differs from 100 by not more than 5 per cent in most cases, whence it follows that dividing $L\sqrt{H_c}/D$ by $\sqrt{B_r}$ shifts the relative position of the abscissas of the points in Fig. 5 by not more than 5 per cent in most cases. Fig. 6 illustrates the results obtained by the second method of plotting.

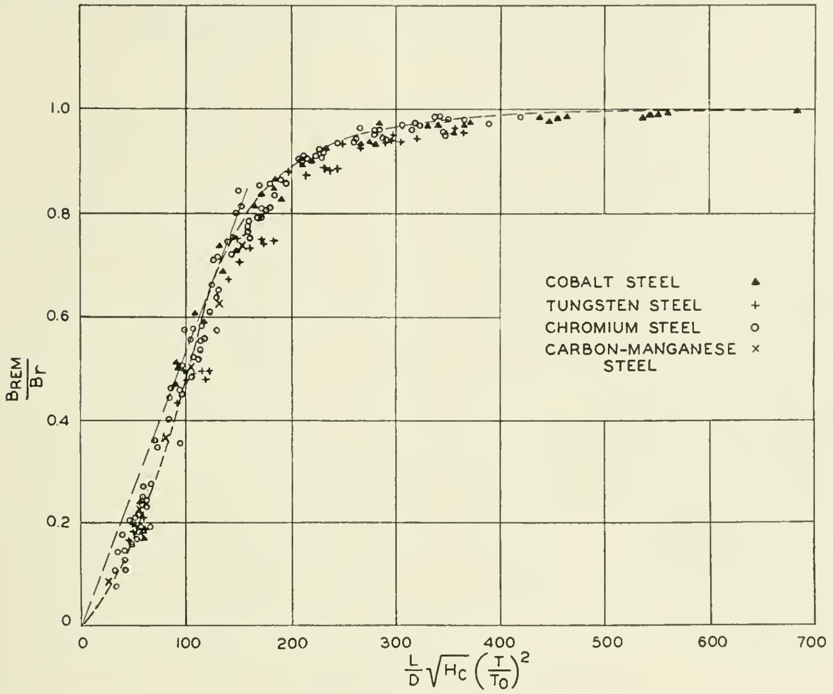


Fig. 5—The same data as in Fig. 4. Showing the correlation between B_{rem}/B_r and $(L/D)\sqrt{H_c}$ for bars of any kind of material, magnetic properties, length or cross-section.

Dimensional considerations favor the use of the quantity $L\sqrt{H_c}/$

$D\sqrt{B_r}$, however, since this quantity, like B_{rem}/B_r, is dimensionless, whereas $L\sqrt{H_c}/D$ is not. For this reason it is felt that the method of plotting shown in Fig. 6 is to be preferred to that of Fig. 5 although the correlation obtained in Fig. 5 appears to be as good as that shown in Fig. 6. Further, the necessity for injecting the variable of temperature into the picture is done away with.

Inasmuch as the data from a large number of bars of widely different compositions, magnetic properties, dimensions and heat treatment are all found to cluster quite closely along a single curve, it is felt that it is legitimate to use the curve as a basis for some generalizations, even though the curve was arrived at largely by empirical processes.

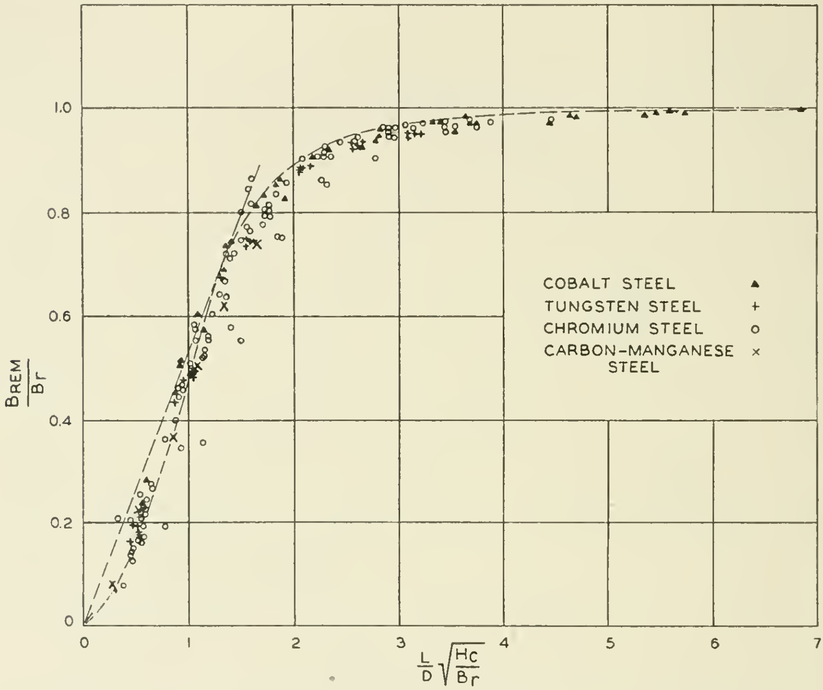


Fig 6.—The data of Fig. 4 with B_{rem}/B_r plotted against $L\sqrt{H_c}/D\sqrt{B_r}$.

The type of curve that is obtained when B_{rem}/B_r is plotted against $L\sqrt{H_c}/D\sqrt{B_r}$ indicates that other factors being constant, the remanence of a magnet is roughly proportional to $\sqrt{B_r}$ and, for large values of $L\sqrt{H_c}/D\sqrt{B_r}$, in other words, for large values of dimension ratio or coercive force, the remanence is practically equal to B_r. This fact is of interest from a design standpoint.

It was to be expected, of course, that the values of B_{rem} would approach B_r as a limit for high values of length of bar, since the end effect diminishes as the length increases and the condition of an infinitely long bar or a closed ring is simulated. The functional relationship which the quantity $L\sqrt{H_c}/D\sqrt{B_r}$ bears to the ration B_{rem}/B_r is not known. A fairly good fit of the observed data is given by the expressions:

$$r = \frac{q^{1.5}}{2.15} \qquad q = O \text{ to } q = 1.25 \text{ and}$$

$$r = e^{-\frac{0.8}{q^{2.75}}} \qquad q = 1.25 \text{ to } q = \infty$$

in which $r = B_{rem}/B_r$ and $q = L\sqrt{H_c/D}\sqrt{B_r}$, but aside from the direct proportionality of B_{rem} and B_r, the equations appear to have little meaning. However, they do indicate that for a given dimension ratio, there is a practical upper limit to H_c, beyond which very large increases in H_c are necessary to produce small increases in the ratio of B_{rem} to B_r. Considerable interest attaches to the fact that the dimension ratio L/D and $\sqrt{H_c}/\sqrt{B_r}$ are of equal weight in affecting the remanence of a magnet. It is also worth noting that the value of B_{rem} is independent of the contour of the cross-sectional area. It is possible that this would not hold for dimension ratios less than one, but it does appear to hold for dimension ratios of practical importance. As shown in Fig. 5, a line drawn through the origin and tangent to the dotted curve of that figure has its point of tangency at values of B_{rem}/B_r and $L\sqrt{H_c}/D\sqrt{B_r}$ of approximately 0.65 and 1.25 respectively. It will be shown later that this is the point of maximum efficiency, *i.e.*, the point at which are obtained the highest values of B_{rem} or external magnetic energy per unit volume of steel.

It follows, if this is true, that magnets should be designed so that $L\sqrt{H_c}/D\sqrt{B_r} = 1.25$, and with this as a basis, the nomogram of Fig. 7 was laid out. The use of this chart to design a magnet for maximum efficiency, is illustrated by the dotted lines of the figure. In the case shown a total flux in the magnet of 3,000 maxwells is desired, and a steel with a B_r of 10,000 gauss and a coercive force of 54 oersteds is assumed. The correct values of A and L are found to be 0.461 cm.^{2} and 13 cm., respectively, and the dimension ratio is 17.

It should be borne in mind throughout this discussion that Figs. 4, 5, 6 and 7 apply directly only to straight bar magnets. Formed magnets with short air-gaps with or without pole pieces, will have higher remanence values than are shown by the curve of Fig. 6. This

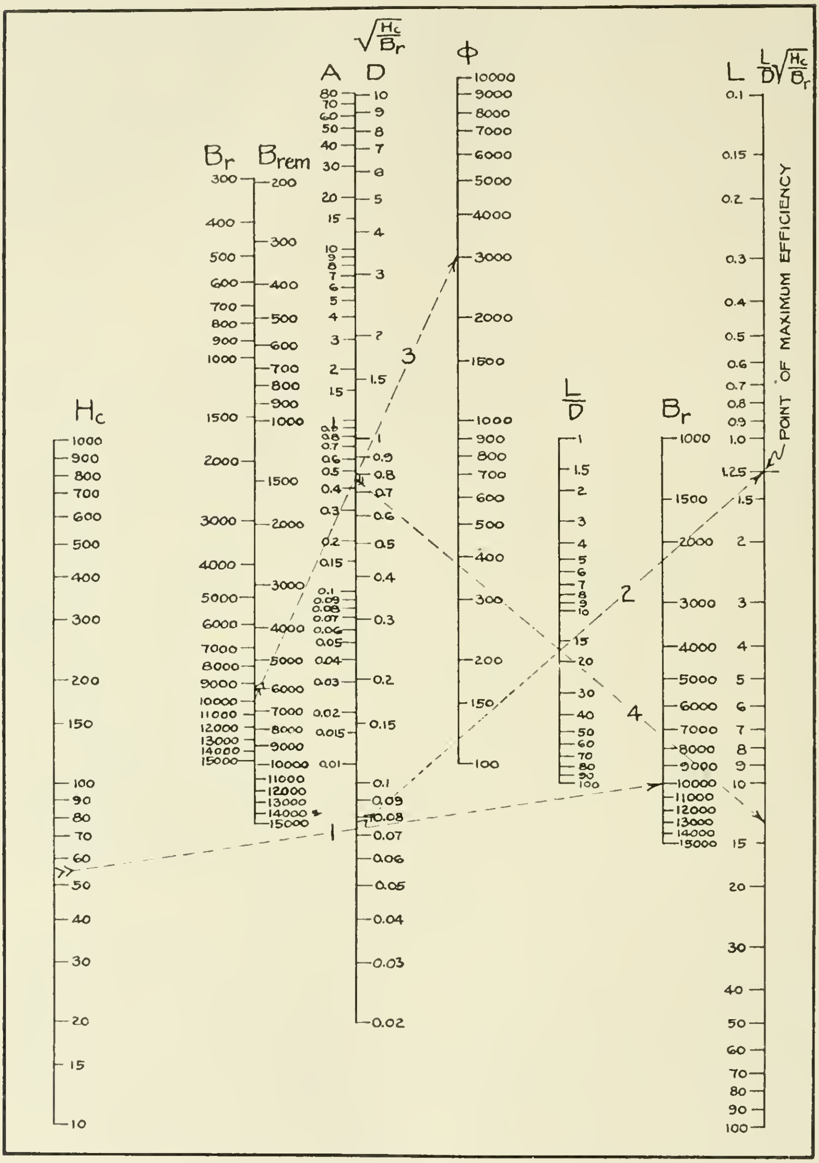


Fig. 7—Nomogram chart for permanent magnet design for maximum efficiency. Based on the curves of Figs. 6 and 10.

is because, compared with straight magnets, their effective lengths are greater than their actual lengths. The difference between effective and actual lengths is negligible, however, except for magnets whose poles are brought quite close together, either actually or by means of pole pieces of soft iron. Watt-hour meter damping magnets are good examples of magnets whose poles are brought thus closely together without the aid of pole pieces.

On the other hand, the general conclusions drawn from the curve of Fig. 6 are applicable to magnets of any shape if L is assumed to stand for the effective length of the magnet. Fig. 7 also is perfectly general on the same basis. By effective length is meant the length of a straight bar magnet of equal cross-section and magnetic properties, having the same remanance as the given magnet.

CORRELATION BETWEEN STRAIGHT AND BENT MAGNETS

Subsequent to the establishment of the relationships shown in Fig. 6, a limited amount of data were secured which have a bearing upon the effect on B_{rem} of departure from the form of a straight bar magnet.

To get these data, one ¼ by ¼ in. rod each of cobalt steel and tungsten steel were formed into rings of 2-in. inside diameter, with the ends touching. Companion test bars were cut adjacent to each end of the rods used in forming the rings. All three pieces of each kind of steel received as nearly identical heat treatments as possible, even to heating and cooling the straight test bars at the time of forming the rings. After hardening, demagnetization curves were obtained on each of the two test bars of tungsten steel and of cobalt steel. Then data for B_{rem} *vs.* length curves were obtained from both the test bars and rings by the process of cutting off the ends and remeasuring B_{rem} already described.

The demagnetization curves and the B_{rem} *vs.* length curves, for the two straight cobalt steel bars are practically identical. Inasmuch as these bars were cut adjacent to the ends of the bar from which the ring was formed, and heat treated with it, it is safe to assume that the material in all three samples is alike magnetically, and that the demagnetization curve for the ring is the same as for the straight bars. The same conclusions hold for the tungsten steel, as shown in Fig. 9.

It follows then that it is valid to compare the B_{rem} *vs.* length curves for the straight bars with those for the rings, and conclude that all differences in the curves are due to differences in shape. It is interesting to note that the ring for cobalt steel yields the same curve for B_{rem} *vs.* length as the straight bar, but that, in the case of the tungsten steel, the proximity of the ends of the ring has an appreciable

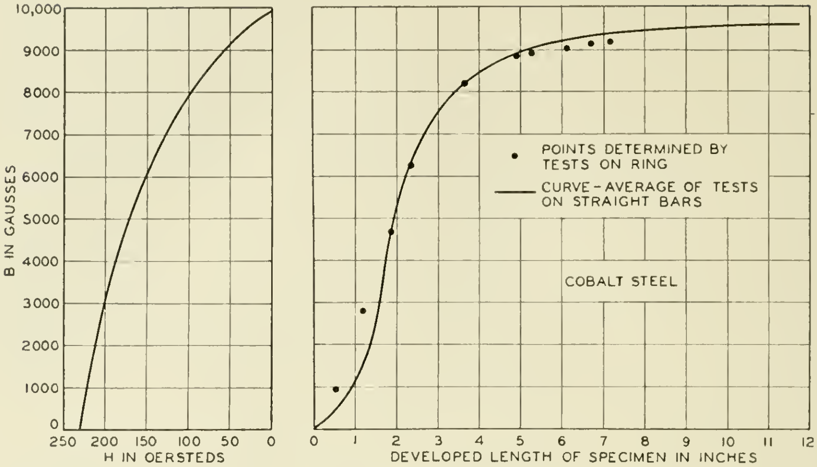


Fig. 8—Comparison of B_{rem} *vs.* length curves for straight bar and 2-in. I.D. ring of ¼-in. x ¼-in. cobalt steel in which B_r = 9,980 gauss and H_c = 230 oersteds.

effect on the value of B_{rem}, causing the curve for B_{rem} *vs.* length to turn upwards for those lengths corresponding to short air gaps.

Additional data along these lines would be very useful, and should include measurements on magnets with pole structures having well defined air gaps.

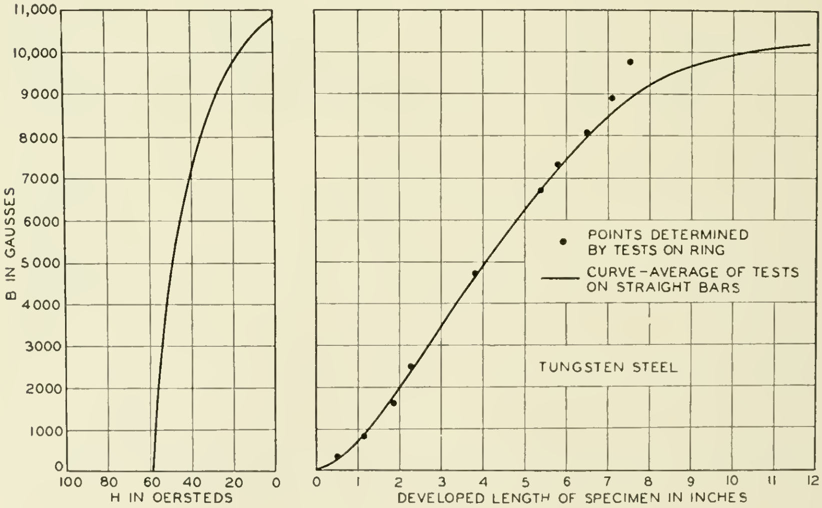


Fig. 9—Comparison of B_{rem} *vs.* length curves for straight bar and 2-in. I.D. ring of ¼-in. x ¼-in. tungsten steel in which B_r = 10,830 gauss and H_c = 59 oersteds.

CRITERIA OF MAGNETIC QUALITY OF PERMANENT MAGNET STEELS

A number of quantities determinable by magnetic measurement have been proposed at various times by different investigators for use as criteria of magnetic quality of magnet steels. Among these quantities may be listed $(BH)_{max}$⁽¹⁾, the maximum product of the coordinates of the demagnetization curve; and the quantities B_rH_c⁽²⁾, B_r/B_{max}, B_rH_c/B_{max} and B_r/H_c⁽³⁾. High values of all of the proposed criteria except the last were considered desirable.

The form B_r/H_c has decided disadvantages, since a low value of the criterion is obtained when B_r is small as well as when H_c is large, and low values of B_r are distinctly undesirable in view of the fact that B_{rem} is directly proportional to B_r. The ratio B_r/B_{max} was found to be practically constant for some types of steel, regardless of the value of H_c, and thus was of no value as a criterion. The quantity $B_r\ H_c/B_{max}$ would accordingly amount to a constant times H_c and would give no actual weight to the value of B_r. These three criteria would therefore fail either to indicate a choice between different types of magnet steel or to aid in the selection of an optimum heat treatment for a given kind of steel.

This leaves the quantities $(BH)_{max}$ and $B_r\ H_c$ as the only ones not obviously faulty. The first was proposed by S. Evershed in a noteworthy paper in the *Journal* of the Institution of Electrical Engineers for September 1920, in which he showed that the quantity $(BH)_{max}$ is a measure of the maximum amount of external magnetic energy which can be supported per unit volume of a given magnet steel, and that this in turn defines the term "magnetic quality" as applied to magnet steel.

Evershed's derivation of his criterion of magnetic quality, given in the article noted above, is complete and convincing, but it was thought desirable to obtain experimental verification. The data from the bars of Fig. 4 were used to this end. In Fig. 10 are shown a number of curves each corresponding to a different type of magnet steel. The demagnetization curves for the bars are shown plotted in the usual way, and to the right of these are plotted the curves of (BH) *vs.* B. To the right of these are plotted further the experimentally determined curves of B_{rem} *vs.* length for the bars in question.

Fig. 10 shows that if tangents to the curves of B_{rem} *vs.* length are drawn through the origin the point of tangency in each case indicates the value of B_{rem} and length at which the ratio of B_{rem} per unit length or volume of steel is a maximum, and it is clearly demonstrated that in each case these values of B_{rem} coincide closely with the values of B for which the product (BH) is a maximum. The writer feels that this constitutes a sufficiently convincing verification of the correctness of Evershed's criterion.

The chief objection to the use of Evershed's criterion is that a number of points on the demagnetization curve for a piece of magnet steel must be known before the numerical value of the product $(BH)_{max}$ can be determined. The quantity $B_r H_c$ is more easily obtained than $(BH)_{max}$ but its validity as a criterion cannot be established from theoretical considerations. Accordingly, the values of $(BH)_{max}$ and $B_r H_c$ were determined by actual measurement for a large number of bars of different kinds of magnet steel. In Fig. 11 are shown the values of $B_r H_c$ plotted against the corresponding values of $(BH)_{max}$.

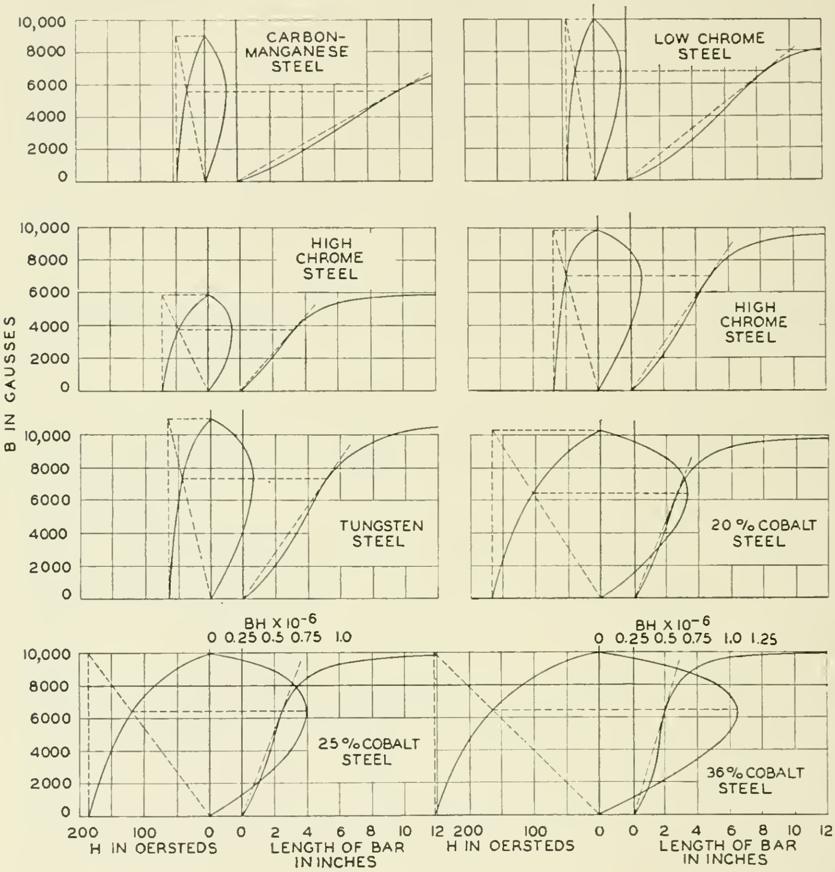


Fig. 10—Curves showing experimental verification of validity of Evershed's criterion for magnetic quality of permanent magnet steels.

It is felt that these points lie closely enough along the straight line in that figure to warrant the use of $B_r H_c$ as a criterion of magnetic quality of magnet steel for all practical purposes. In cases where very accurate comparisons are required, it may be necessary to go to the greater trouble of determining $(BH)_{max}$.

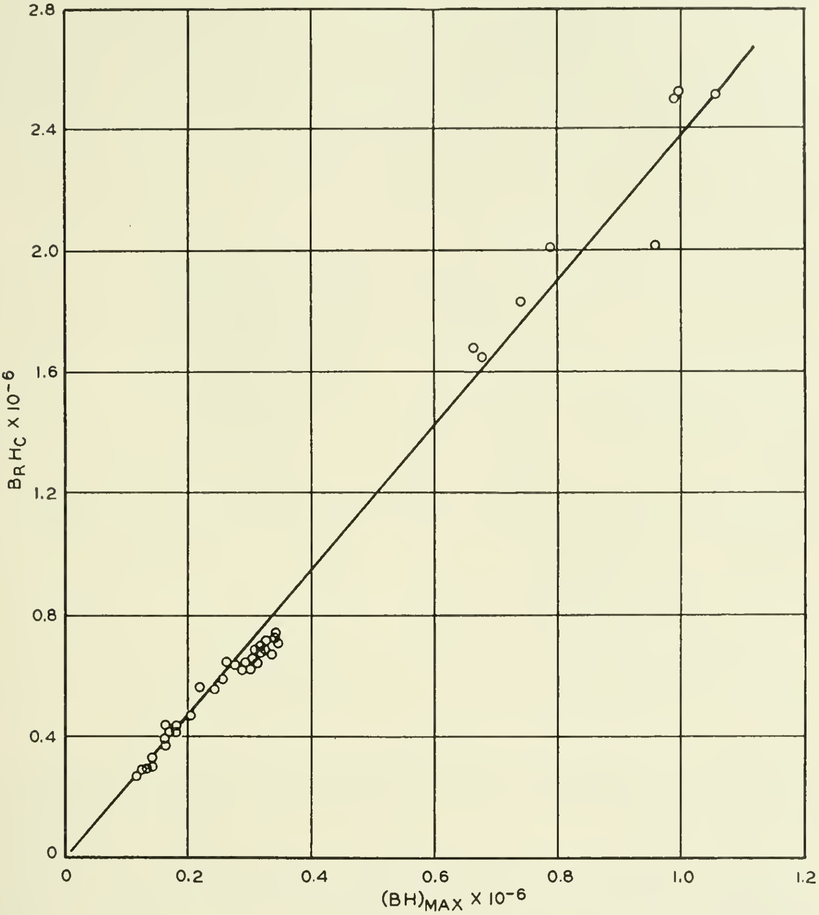


Fig. 11—Relation between the quantity $B_r \times H_c$ and Evershed's criterion, $(BH)_{max}$.

Referring to a previous mention of optimum hardening temperatures, it will be understandable now to state that the optimum hardening temperature for a given type of steel is that temperature at which a maximum value of $(BH)_{max}$ is obtained, or practically speaking, a maximum value of $B_r H_c$.

Fit of Demagnetization Curve by Hyperbola

It will be noted that the points on the demagnetization curves for which the product (BH) is a maximum are given quite accurately in each case by the intersection of the demagnetization curve with a line through the origin having the slope B_r/H_c. That this should be so follows from the fact that any demagnetization curve for magnet steel may be closely approximated by a rectangular hyperbola whose equation is $B = a - k/(H + b)$ in which a, b, and k are parameters of each particular curve. It is a mathematical property of the foregoing rectangular hyperbola that the coordinates whose product is a maximum are located by a line through the origin having a slope equal to the ratio of the intercepts of the hyperbola. This property of the hyperbola is a property of the demagnetization curves for magnet steel in so far as these curves can be closely fitted by appropriate hyperbolas. That the fit is quite good in all cases is shown by Fig. 12 in which hyperbolas calculated for each demagnetization curve are shown by dots and the observed points by a full line. The graphical method of Fig. 10 for determining the point on each demagnetization curve the product of whose coordinates is a maximum is more accurate than the method of plotting the curve of (BH) *vs.* B, because the latter curve is usually quite flat-topped and its maximum is hard to locate exactly. The fact that the curve is flat-topped explains why many methods of magnet design give good results. It is because about equal efficiencies are obtained in any case over a fairly wide range of values of B_{rem}.

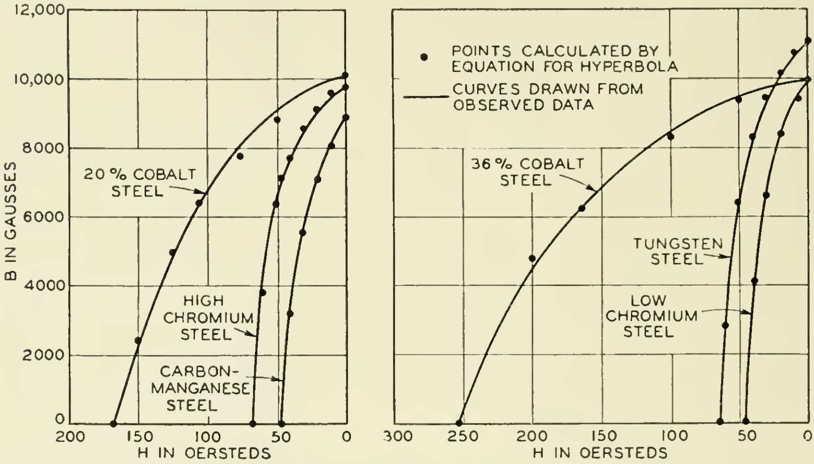


Fig. 12—Curves showing approximation of actual demagnetization curves by points on the general hyperbola, $B = a - k/(H + b)$.

A further application of the idea that the demagnetization curve can be represented by a hyperbola is made by Mr. E. A. Watson.[4] Watson applies a graphical construction for a hyperbola to the demagnetization curve and extends the construction into the region of positive values of H, thus providing a means of extrapolating the hysteresis loop to the saturation value of B. It has been the writer's experience that Watson's graphical construction cannot be legitimately extended into the first quadrant because the curve thus predetermined almost always lies considerably under the true curve, and yields too low a value for the saturation flux density. Fig. 13 illustrates the case in point.

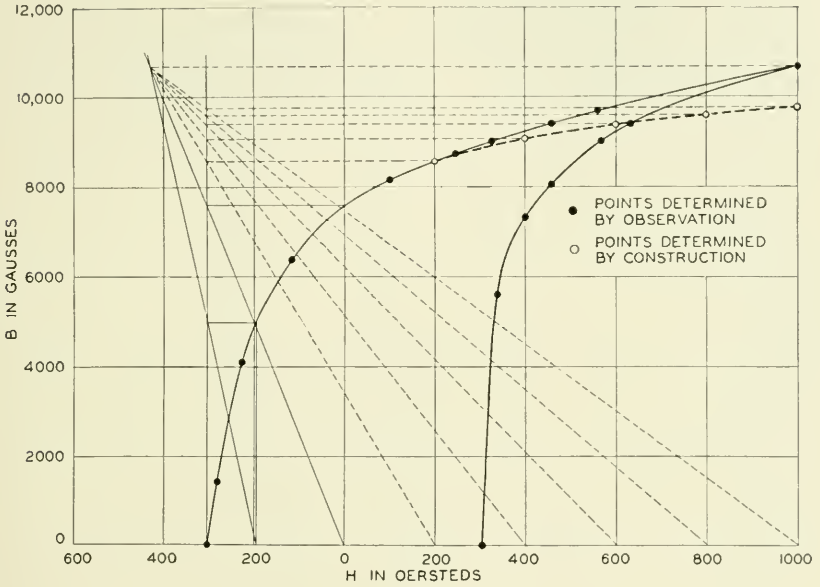


Fig. 13—Comparison of graphical construction of the descending portion of the hysteresis loop with actual measured data. Note the lack of agreement for positive values of H.

If, instead of potting B *vs.* H for the points of the demagnetization curve, we plot B/B_r *vs.* H/H_c, it is found that all test specimens yield points fitting fairly well the single curve represented by the equation $y = A(x + 1)/(x + A)$, in which $y = B/B_r$, $x = H/H_c$ and $A = \sqrt{2}$ approximately. It is possible to derive from this equation the relationship that $(BH)_{max} = 0.423\ B_r\ H_c$ which checks very well with the line drawn in Fig. 11. It can also be seen from the equation $y = A(x + 1)/(x + A)$ that the representation of the demagnetization curve by a hyperbola cannot be extrapolated to give the correct value for B_s since for values of H approaching infinity, the equation yields the result that $y = B_s/B_r = A = \sqrt{2}$ whence $B_s = \sqrt{2}\, B_r$, while values of B in excess of $\sqrt{2}\, B_r$ are commonly encountered at ordinary values of H.

Characteristic Curves for Magnet Steel

In Fig. 14 are shown demagnetization curves and (BH) curves for the various types of magnet steel with which the writer has had experience. It should be remembered that while these curves are more or less typical of the steels in question, there is no one curve which can be given as the curve for the material. There will be as many curves as there are samples tested, because magnetic characteristics of a piece of magnet steel, or any other magnetic material depend upon the whole previous thermal history of the sample from the time of melting to the time of testing, the amount of hot and cold working the sample has received, and the conditions of stress in the material

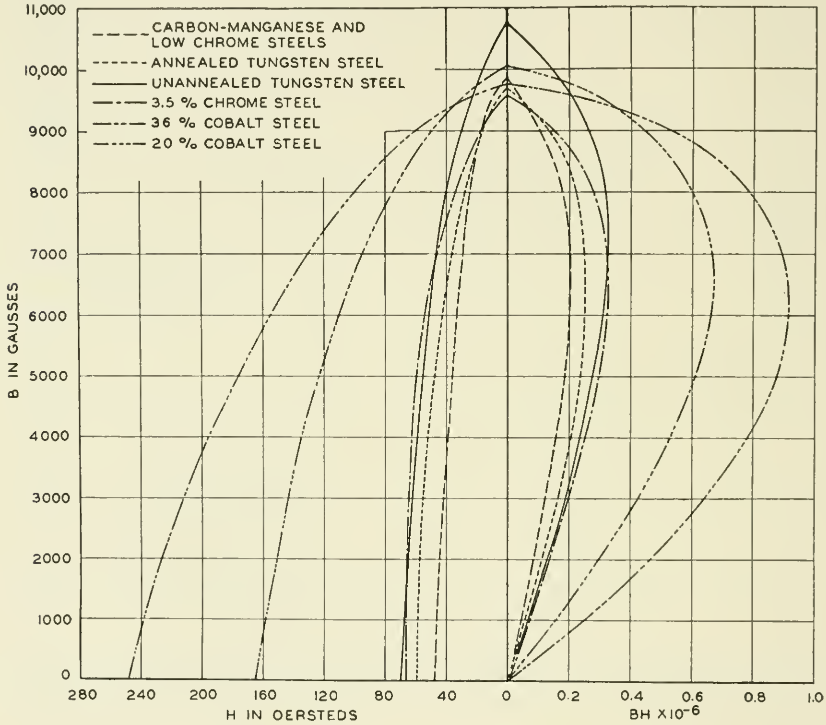


Fig. 14—Typical demagnetization curves and curves of BH vs. B for various kinds of magnet steel.

at the time of testing. The relation of all of these variables to the remanence of a permanent magnet, fortunately, is focussed, as it were, in the values of B_r and H_c for the material, and the relation of these variables to B_{rem} is as shown by Fig. 6.

REFERENCES

1. "Permanent Magnets in Theory and Practise," Evershed, S., *Journ. Inst. Elect. Eng.* **58**, 780 (1920); **63**, 725 (1925).
2. "Ein neues Material für permanente Magnete," Gumlich, E., *Elektrot. Zeits.* **44**, 147 (1923).
3. "Magnetic Habits of Alloy Steels," Mathews, J. A., *Proc. Amer. Soc. Testing Mat.* **14**, 50 (1914).
4. In discussion, *loc. cit.*(1), *Journ. Inst. Elect. Eng.* **58**, 829 (1920), also *ibid.* **61**, 641 (1923).
5. "Calcul des appareils magnétiques, Diagramme géneralé," Picou, R. V., *Rev. gén. dé l'élect.*, **22**, 259 (1928).
6. "Une solution sans fiction du probléme de l'attraction magnétique," Lehmann, T., *Rev. gén. dé l'élect.* **20**, 433 (1926).
7. "An Improved Permeameter for Testing Magnet Steel," Babbitt, B. J., *Jour. Opt. Soc. Amer.* **17**, 47 (1928).

A Method of Measuring Acoustic Impedance *

By P. B. FLANDERS

An apparatus is described whereby acoustic impedances may be measured in terms of a known acoustic impedance and the complex ratios of two electrical potentiometer readings to a third. As a known impedance, there is chosen the reactance of a closed tube of uniform bore which is an eighth wave-length long. The electrical readings are obtained by balancing the amplified output of a condenser transmitter against the electrical input of the source of sound. The condenser transmitter picks up the acoustic pressure at the junction of the sound-source and the attached impedance. A balance is made for each of three successively attached impedances: (1) a closed tube an eighth wave-length long, (2) a rigid closure of the sound-source, and (3) the impedance to be measured. The unknown acoustic impedance Z is then calculated in terms of the known acoustic impedance Z_0 by means of the equation $Z = Z_0 \frac{z_1 - z_2}{z_3 - z_2}$, where z_1, z_2 and z_3 are, respectively, the three electrical impedance settings of the potentiometer. As indicated by this equation, the constants of the electrical circuit are involved only as ratios, so that the response characteristics of the source of the sound, condenser transmitter and amplifiers (provided they are invariable) do not affect the measurement.

Illustrations are given of impedance measurements on a closed tube of uniform bore, a conical horn, an exponential horn, an " infinite " tube, and a hole in an " infinite " wall.

THE progress in acoustics during the past few years has caused acoustic impedance measurement to have the same relative importance that impedance measurement in electrical work has had for many years. The concept of acoustic impedance is derived from the analogy [1] that exists between electrical and acoustic devices, as shown by the analogous differential equations describing their action. Acoustic impedance is usually defined as the complex ratio of pressure to volume velocity (or flux) but it is sometimes more convenient to deal with ratios of pressure to linear velocity or force to linear velocity. The magnitudes of these are interrelated, of course, by powers of the area involved.

The earliest efforts to measure acoustic impedance seem to have been made by Kennelly and Kurokawa.[2] In their method, electrical measurements were made of the motional impedance of a telephone receiver, with and without an attached acoustic impedance. Except for frequencies near resonance, the method was inaccurate because the acoustic impedance was associated with a relatively large mechanical impedance.

* Presented before Acous. Soc. Amer., New York City, May 3, 1932.

[1] This analogy was first pointed out by A. G. Webster in *Nat. Acad. of Science*, **5,** 275 (1919).

[2] *Proc. Am. Ac. Arts and Sc.*, **56,** 1 (1921).

Later, a direct method was described by Stewart[3] who measured the change in acoustic transmission through a long uniform tube when the unknown impedance was inserted as a branch.

The apparatus to be described in this paper measures acoustic impedance directly in terms of a known acoustic impedance and three balance readings of an electrical potentiometer. The only assumptions involved in the method are that the elements of the apparatus be invariable during a measurement, and that the value of the comparison acoustic impedance be known accurately.

APPARATUS

Fig. 1 shows the general arrangement of the apparatus. An oscillator feeds electrical energy into a loud speaker where a portion is converted into acoustic energy which travels along the tube and into an attached impedance. A canal about 0.06 inches in diameter picks up the sound pressure at the junction of the tube and the attached impedance, and passes it along to a small condenser transmitter. A corresponding voltage, generated by the transmitter, is amplified and the current output of the amplifier passed through a variable resistance in series with the primary of a variable mutual inductance.

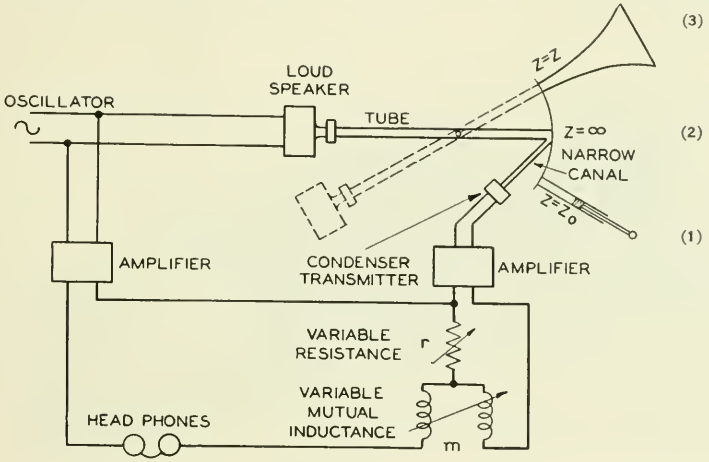


Fig. 1—Schematic circuit of acoustic impedance measuring apparatus.

The same oscillator also feeds energy through a second amplifier at the output of which the voltage is balanced (by the null method) against the voltage drop across the variable resistance and the secondary of the mutual inductance.

---

[3] *Phys. Rev.*, **28**, 1038 (1926).

At the end of the tube from the loud speaker are three different impedances. One is the reactance offered by a closed tube of uniform bore. The closure is formed by a well-fitting plunger whose position in the tube may be adjusted. The second is the infinite impedance offered by a rigid wall closing the end of the tube from the loud speaker. The third is the impedance to be measured. All three impedance elements are fixed in position. The loud speaker, tube, condenser transmitter and associated amplifiers are, however, mounted together on a carriage which can be rotated so as to bring any one of these impedances into alignment with the tube. For brevity, reference to these three positions will hereinafter be to positions 1, 2, and 3.

For any one frequency a balance is obtained for each of the three positions. These three electrical readings and the reactance value of the closed tube are sufficient to determine the impedance being measured.

A photograph of the apparatus is shown in Fig. 2.

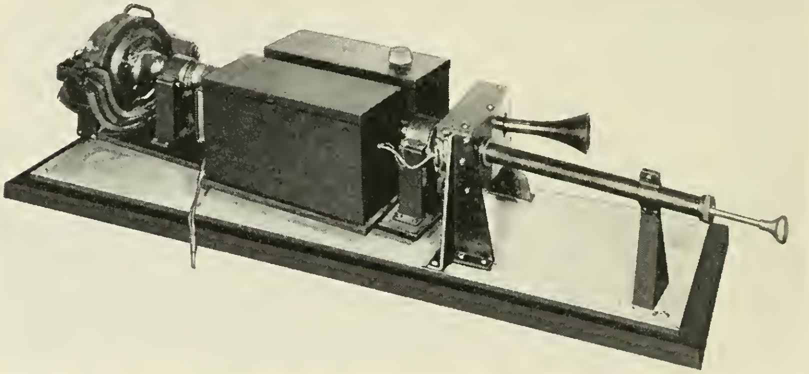


Fig. 2—View of acoustic impedance equipment, showing loud speaker, tubes, condenser transmitter amplifiers and small horn in position for measurement.

THEORY

Thevenin's theorem [4] states that, in an invariable electrical network, the current in any branch is equal to the current that would flow in a simple series circuit composed of an electromotive force and two impedances. The electromotive force is the voltage that would obtain at the branch terminals on open circuit. The impedances are the impedance at the terminals looking back into the source of power, and the impedance of the branch.

[4] K. S. Johnson's "Transmission Circuits for Telephone Communication," Ch VIII.

Since the differential equations of acoustics are analogous to those of electrical lines and networks, the theorem may be applied to the action of this apparatus with a considerable saving in labor.[5]

By Thevenin's theorem then, the tube, loud speaker, oscillator, etc. may be replaced by one pressure and one impedance. The pressure is the "open-circuit" pressure at the end of the tube, or in other words the pressure that would be exerted on a rigid wall if placed there. The impedance is the complex ratio of pressure to velocity at the end of the tube which would exist if acoustic energy were sent into it toward the loud speaker, the oscillator being shut off. In electrical terms, this would be called the impedance looking into the source. The velocity or acoustic current that flows into an impedance attached to the end is then the current that would flow in an analogous circuit composed of this vibromotive force or pressure and the two impedances in series. This impedance diagram is given in Fig. 3.

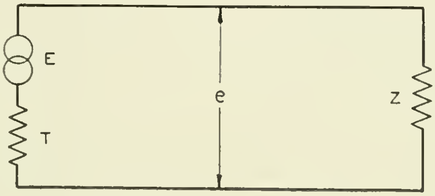


Fig. 3—Impedance diagram for Thevenin's theorem.

E is the open-circuit voltage or pressure, T the impedance looking into the source of sound at the junction, and Z the attached impedance. The pressure e at the junction of T and Z is, of course, the velocity-current $\frac{E}{T+Z}$ in the loop, multiplied by Z. The three equations for three values of attached impedance are

$$Z = Z_0, \qquad e_1 = \frac{EZ_0}{T+Z_0},$$

$$Z = \infty, \qquad e_2 = E,$$

$$Z = Z, \qquad e_3 = \frac{EZ}{T+Z}.$$

The two unknown quantities, E and T, can be eliminated giving one equation

$$\frac{Z}{Z_0} = \frac{\dfrac{e_2}{e_1} - 1}{\dfrac{e_2}{e_3} - 1},$$

---

[5] A more direct proof of Thevenin's theorem as applied to acoustics is given by W. P. Mason in *B. S. T. J.*, 6, 291 (1927).

whereby Z may be calculated in terms of Z_0 and two ratios of pressures at the junction.

Referring now to Fig. 1 it will be seen that the current through the resistance and mutual primary is proportional to the pressure at the junction. The drop in voltage across the secondary and the resistance is equal in magnitude and opposite in phase to a voltage proportional to E, when no current passes through the head-phones. If k signifies the circuit constant and if z be the impedance value of the resistance and the mutual inductance, then $ez = kE$; and the above equation becomes

$$\frac{Z}{Z_0} = \frac{\dfrac{z_1}{z_2} - 1}{\dfrac{z_3}{z_2} - 1} = \frac{z_1 - z_2}{z_3 - z_2} = \frac{(r_1 - r_2) + j\omega(m_1 - m_2)}{(r_3 - r_2) + j\omega(m_3 - m_2)},$$

where r is the resistance component of z and m is the mutual inductance.

The reactance of a closed tube of uniform bore whose length is one-eighth the wave length of sound for the measuring frequency is chosen as the known impedance. If dissipation in the tube be neglected, the impedance is readily calculated [6] to be a pure negative reactance of 41 mechanical ohms per square centimeter [7] at a temperature of 20° C. This value is chosen because it is of the same order of magnitude as most acoustic impedances. By mechanical ohms per square centimeter is meant the complex ratio of pressure to the linear velocity of the air. The justification for assuming negligible dissipation will be apparent when measurements made on a closed tube, several wave-lengths long, are described.

In making measurements, the three impedance values necessary for balance are read for the three impedance conditions in the 2-1-3 or 2-3-1 order. Afterwards, as a check to ensure that the circuit constant has not changed during the measurement, condition 2 is measured again. This series of four measurements is repeated for each frequency.

Application

Fig. 4 shows the results of measuring the reactance of a closed tube. The tube was 2.4 inches long and 0.7 inch in diameter. The comparison impedance was the calculated reactance of this same tube in the one-eighth wave-length condition, assuming no dissipation. The impedance was also calculated,[8] taking into account viscosity and losses through heat conduction, for frequencies near the half wavelength anti-resonance, where dissipative effects are most pronounced. It will be seen from Fig. 5 that there is a close agreement between the theoretical curve and the measured points. It seems reasonable, therefore, to assume that the value chosen for the comparison impedance is quite accurate.

---

[6] See I. B. Crandall's "Theory of Vibrating Systems and Sound," p. 104.

[7] See definitions 8007 and 8011 in "Standardization Report of I. R. E." in "Year Book of I. R. E.," 1931.

[8] See Rayleigh, "Theory of Sound," Vol. II, pp 318 and 325.

Fig. 6 shows the impedance of a conical horn and Fig. 7 that of an exponential horn. In both cases, the mouth of the horn projected through a window into open air, so as to minimize reflection effects from external objects. Reflection effects from the mouth, where there is a change in impedance, are present, however, and these appear as oscillations of the impedance about a mean which is the characteristic impedance of the horn. By characteristic impedance is meant the impedance that would obtain looking into the throat of the horn were it infinite in length.

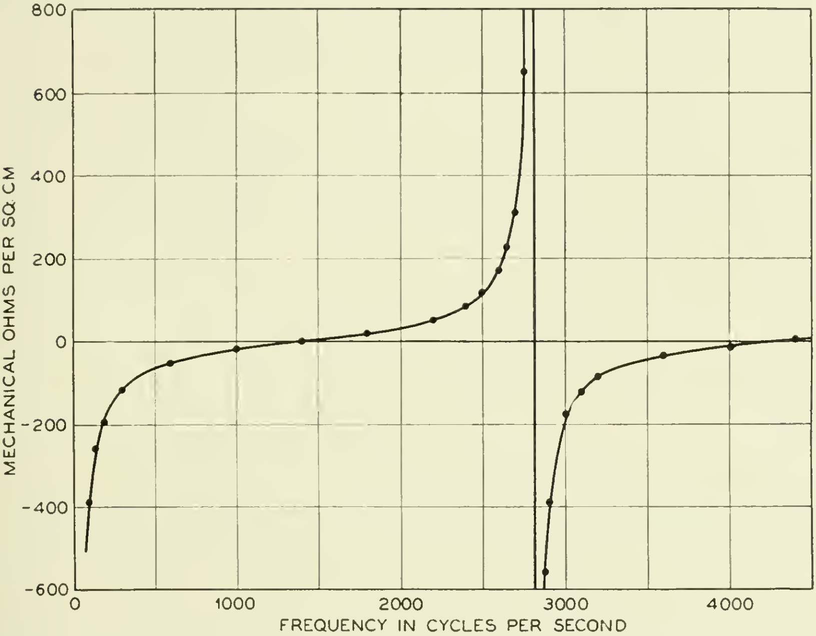


Fig. 4—Acoustic reactance of closed cylindrical tube, 2.4 inches long and 0.7 inch in diameter.

Fig. 8 is the impedance of an "infinite" tube. The tube was actually 112 feet long and coiled into a helix. At low frequencies, where the dissipative losses are small, reflection effects from the open end are

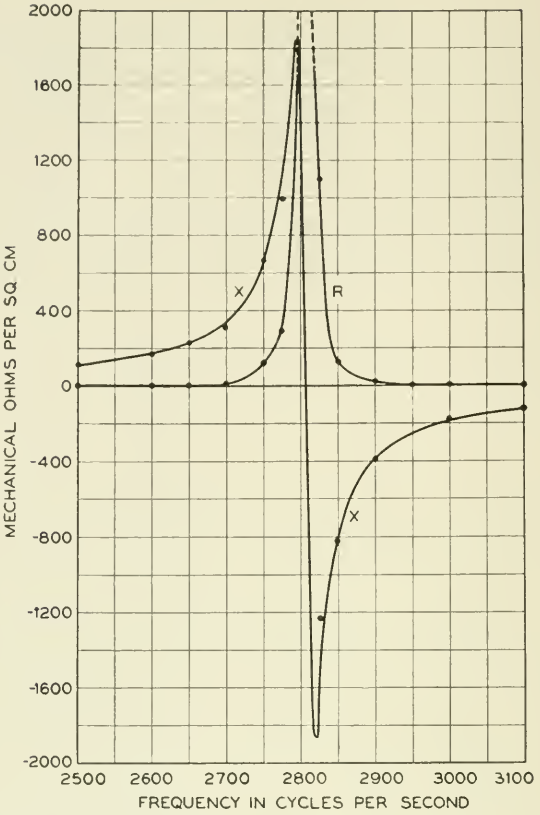


Fig. 5—Acoustic impedance of closed cylindrical tube, 2.4 inches long and 0.7 inch in diameter, showing agreement between measured and calculated values.

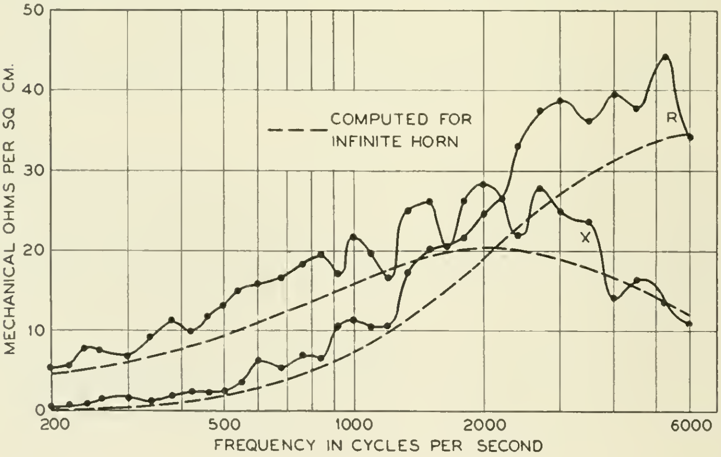


Fig. 6—Acoustic impedance of 38 inch conical horn, having end diameters of 0.7 inch and 28 inches.

observed as oscillations of the impedance at about 5-cycle intervals. An examination of the measurements in this oscillatory region (Fig. 9) will make evident the precision of the apparatus.

Fig. 10 shows the radiation impedance of a hole, 0.7 inch in diameter and surrounded by a flange which approximates an infinite wall for

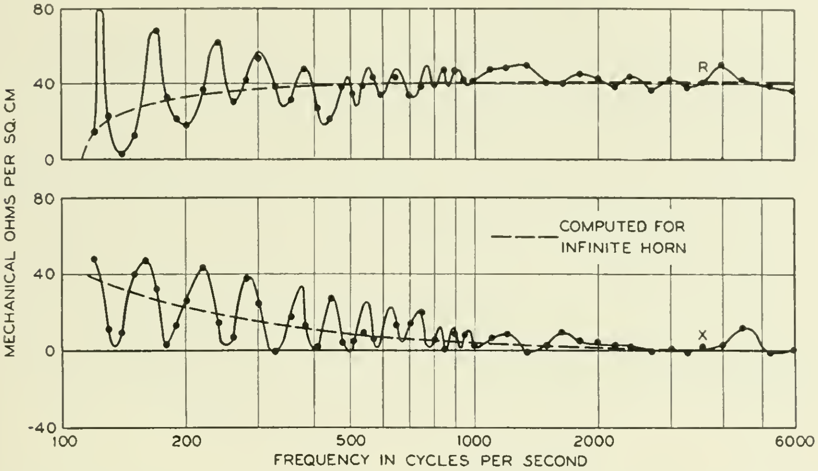


Fig. 7—Acoustic impedance of 6 foot exponential horn, having end diameters of 0.7 inch and 30 inches.

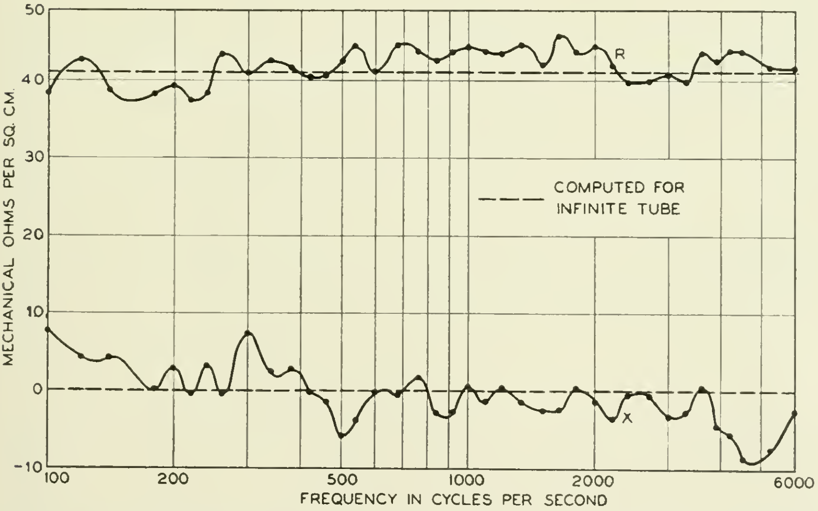


Fig. 8—Acoustic impedance of 112 foot open tube, 0.7 inch in diameter, coiled into helix. Measuring frequencies chosen at random.

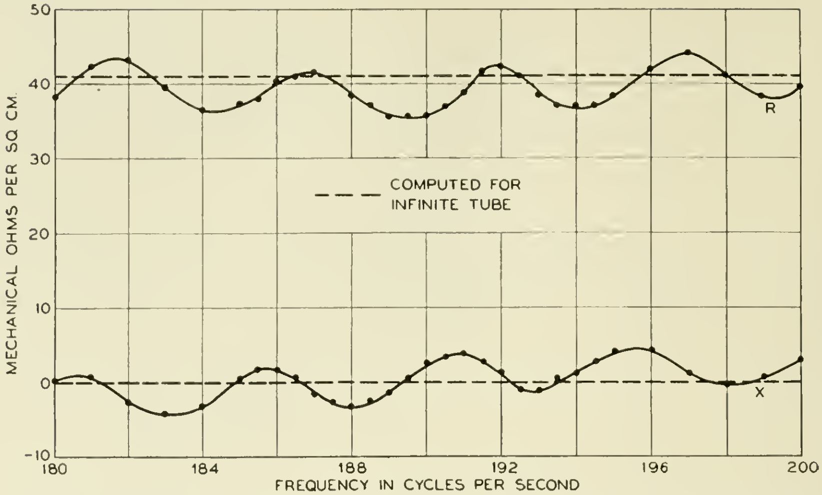


Fig. 9—Acoustic impedance of 112 foot open tube, 0.7 inch in diameter, coiled into helix. Measuring frequencies chosen at half and one cycle intervals to show oscillatory character of impedance.

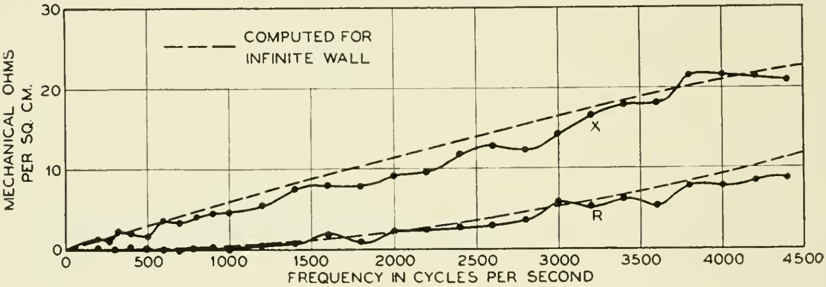


Fig. 10—Acoustic radiation impedance of hole, 0.7 inch in diameter, in flange having diameter of 6 inches.

the frequencies of interest. The dotted lines are the resistance and reactance as calculated by the equations of Rayleigh.[9]

[9] "Theory of Sound," Vol. II, p. 164.

Transmission Lines for Short-Wave Radio Systems*

By E. J. STERBA AND C. B. FELDMAN

The requirements imposed on transmission lines by short-wave radio systems are discussed, and the difference in the requirements for transmitting and receiving purposes is emphasized. Various line types are discussed, particular attention being given to concentric tube lines and balanced two-wire lines. The concentric tube line is particularly valuable in receiving stations where great directional discrimination is involved and low noise and static pick-up is required.

Excellent agreement between calculations and measurements is found for the high-frequency resistance of concentric lines, using the asymptotic skin effect formula of Russell. Other losses in correctly designed concentric tube lines are found to be negligible. Measured losses in two-wire lines are found to be greater than losses predicted by the asymptotic skin effect formula owing, in part, to losses brought about by unbalanced currents.

Practical aspects of line construction such as joints, insulation, and provision for expansion with increasing temperature are discussed.

Some difficulties encountered in transmission line practice, such as losses due to radiation, reflections from irregularities, effects of weather, and spurious couplings between antenna and line are discussed.

I. General Requirements

THE transmission line systems employed for the purpose of transferring energy between radio units and antennas are fundamentally no different from line systems used in power or telephone work. Owing, however, to the high frequencies employed in radio transmission an operating technique differing from that found economical in low-frequency practice is necessary. An important consideration in 60-cycle power practice is that the voltage at the far end of the line be maintained constant irrespective of load variations. At radio frequencies a transmission line may be many wave-lengths long and the reflections from a load other than one equal to the characteristic impedance of the line produce standing waves. Transmission losses in radio-frequency lines are appreciably augmented when the currents and voltages on the line appear in the form of standing waves. The operation of the radio unit connected to the line is sometimes affected by the presence of standing waves.

Induction and cross-talk problems familiar to every telephone engineer are increasingly important as line operation approaches radio frequencies. Owing to the high sensitivity of radio receiving equipment as compared with that of telephone equipment the difficulties arising from static and noise pick-up are more marked at radio frequencies. Spurious radiations from lines operated at radio frequencies may completely destroy the directional characteristics of an antenna and in addition may cause interference to other radio stations. Under certain conditions the spurious power radiated by a transmission line may be an appreciable fraction of that radiated by the antenna connected to the line.

---

* Presented at I. R. E. Convention Pittsburgh, Pa., April 7–9, 1932. Published in *Proc. I. R. E.*, July, 1932.

It follows that although the primary purpose of a transmission line in a radio station is to provide a means for transferring energy between an antenna and the radio unit, a consideration of great importance is the degree of isolation from its associated antenna, from other antennas and lines, and from extraneous sources of signals. This is particularly true in receiving stations where discrimination against undesired signals is oftentimes of greater importance than the over-all sensitivity to the desired signal. Extraneous pick-up on a line to the receiving unit may not only destroy the directional pattern of the antenna but it may also introduce a noise level into the receiver output comparable to the desired signal and so destroy the utility of the apparatus. As compared with transmitters, receivers are generally small units. It is economical to house several units in one building. The lines not only to one receiver but those to other receivers must of necessity be in close proximity. Thus the possibilities for cross-talk with ensuing increase in noise levels, loss of circuit gain, and loss of discrimination are greatly augmented.

Of course, cross-talk possibilities between lines of adjacent transmitters cannot be ignored. Transmitters, however, usually occupy sufficient space so that a desirable degree of line separation is obtained. With the exception of local lightning storms static pick-up is of no great importance. High voltage surges due to lightning may be drained by means of properly placed horn gaps and grounds.

Insulation for high voltages at radio frequencies is an important consideration for the case of lines connected to transmitters. Insulators for balanced open-wire construction may be selected from materials designed for high voltage power transmission. However, insulators for concentric tube lines capable of transmitting several kilowatts of modulated radio-frequency power require special consideration.

The lines commonly employed in radio stations may be divided into four classes: single-wire lines, balanced open-wire lines, multiple-wire lines, and concentric-tube lines.

Single-wire lines are of limited utility owing to the low efficiencies arising from the marked radiation characteristics of such wires. The power radiated by a single-wire line several wave-lengths long may be equal to that radiated by the antenna to which it is connected.[1] In fact, single-wire lines, particularly when terminated, are for certain services desirable radiating elements. Diamond-shaped arrays of such elements are employed in some of the radio facilities of the Bell System.[2]

It is generally appreciated that the power losses due to radiation may be reduced by employing two conductors in a go-and-return circuit, the wires being separated a small fraction of a wave-length. A necessary requirement is that the two wires carry equal currents exactly opposite in phase. Otherwise, there will appear current components which employ the two conductors in parallel. In the latter event the radiation losses ascribed to single-wire conductors occur.

Although there is a very great reduction in radiated power in balanced two-wire lines as compared with single-wire lines there are many practical cases where the radiation from two-wire lines produces cross-talk and loss of signal discrimination. Multiple-wire lines comprising several pairs of conductors in go-and-return circuits may be employed to reduce the undesired radiation couplings. As in the two-wire case, care must be exercised in maintaining the required current amplitudes and phases since otherwise the radiation losses ascribed to single-wire lines may destroy the utility of the multiple-wire system. Multiple-wire lines, of course, reduce static and noise interference.

From the standpoint of isolation an ideal electrical connection between antennas and radio apparatus is approached when one conductor completely encloses the other conductor. A concentric-tube line comprising an outer sheath and an inner conductor is the practical form of this construction. Long transmission lines often pick up a large amount of static and other electrical disturbances. Spurious couplings may introduce these disturbances into the radio circuit. Electrical disturbances so introduced are greatly reduced when the outer sheath of a line may be grounded at frequent intervals. In fact, concentric-tube lines may be buried in the ground.

The effect of weather is a factor which in some instances may determine the type of construction to be employed in radio-frequency lines. It is generally appreciated that rain and sleet storms may materially lower the insulation of a line. The velocity of propagation and characteristic impedance also are affected by a coating of water or sleet upon the wires. Concentric-tube lines may be constructed so as to be weather proof.

This paper will be confined entirely to concentric-tube lines, to balanced two-wire lines, and to the apparatus associated with these two line types.

---

[1] See calculations in the appendix.

[2] E. Bruce, *Proc. I. R. E.*, p. 1406, August, 1931.

II. Concentric-Tube Lines

A shielded line comprising an inner tubular conductor and an outer concentric shield is the form most commonly employed in radio practice. Owing to the circular symmetry of the line the case is capable of rather exact mathematical analysis. At radio frequencies the results are surprisingly simple. This simplicity is very evident for the two important parameters of a transmission line, the propagation constant P and the characteristic impedance Z_0.

The propagation constant of a line is defined by:

$$P = \sqrt{R + j\omega L}\cdot\sqrt{G + j\omega C}, \tag{1}$$

in which

$(R + j\omega L)$ is the complex impedance and

$(G + j\omega C)$ is the complex admittance, both per unit length. It is well known that the propagation constant is a complex number and that at radio frequencies (1) reduces to:[3]

$$P = \alpha + j\beta = \frac{R}{2Z_0} + \frac{GZ_0}{2} + \frac{j2\pi}{\lambda}, \tag{1a}$$

in which R is the resistance and G is the leakage conductance, both per unit length and at the wave-length λ.

The characteristic impedance Z_0 is defined by the ratio:

$$Z_0 = \frac{\sqrt{R + j\omega L}}{\sqrt{G + j\omega C}}\cdot \tag{2}$$

The characteristic impedance also is a complex quantity, but at radio frequencies it is for most practical purposes the real quantity:

$$Z_0 = \sqrt{\frac{L}{C}}\cdot \tag{2a}$$

In the case of concentric-tube lines the expression for the capacity C per unit length is the familiar relation:

$$C = \frac{1}{2\log_e \frac{b}{a}}\ \text{e.s.u.}, \tag{2b}$$

in which a is the outer radius of the inner conductor and b is the inner radius of the outer conductor. The inductance L per unit length may be obtained from an expression derived by Lord Rayleigh upon assuming that the two tubes comprising the line are of negligible thickness. This is permissible because at radio frequencies the conduction of currents is essentially a skin effect. Upon this basis the inductance per unit length of a concentric-tube line becomes:

---

[3] J. A. Fleming, "The Propagation of Electric Currents."

$$L = 2 \log_e \frac{b}{a} \text{ e.m.u.} \tag{2c}$$

Upon substituting (2b) and (2c) into (2a) with proper regard of units a simple expression for the characteristic impedance at radio frequencies is obtained:

$$Z_0 = 138 \log_{10} \frac{b}{a} \text{ ohms.} \tag{2d}$$

The high-frequency resistance of concentric-tube lines has been treated by a number of investigators, notably by A. Russell.[4] The asymptotic formula for resistance as the frequency is increased without limit is:

$$R = \sqrt{\rho \mu f}\left(\frac{1}{a} + \frac{1}{b}\right) 10^{-9} \text{ ohms/cm,} \tag{3}$$

in which:

ρ is the resistivity in e.m.u. (for pure copper ρ is about 1730 e.m.u.),

μ is the magnetic permeability,

f is the frequency, c.p.s.,

a is the outer radius of the inner conductor,

and

b is the inner radius of the outer conductor, the two latter being in centimeters.

It is of interest to note that the wall thickness of the conductor is not involved. At radio frequencies the current is confined to a very thin layer on the outside of the inner conductor and on the inside of the outer conductor.[5] The skin effect is, of course, not so pronounced at low frequencies and more complicated formulas involving wall thickness must be employed.

Some typical experimental data are submitted to show that for frequencies higher than one megacycle and for several practical line constructions the foregoing equation (3) holds with a very useful degree of accuracy. The physical dimensions and construction details of

[4] A. Russell, *Phil. Mag.*, April 1909; and "Alternating Currents," Vol. I, p. 222, 1914, Cambridge Press.

[5] Frequency is not the sole criterion, resistivity, wall thickness, and diameter also being involved.

the lines for which the observations were made appear in Fig. 1. With the exception of one rubber insulated line all inner conductors were supported on porcelain insulators. The latter were attached to the inner conductor by means of spring clips, extruded metal ears or by means of soldered rings. Some measurements were made on lines assembled with soldered joints and some on lines connected by means of pipe unions with miniature plug and jack connections for the inner pipe. Various line lengths were employed. Most of the observations comprised measurements of the quantity $(R/2 + GZ_0^2/2)$. The measurement procedure will be described later.

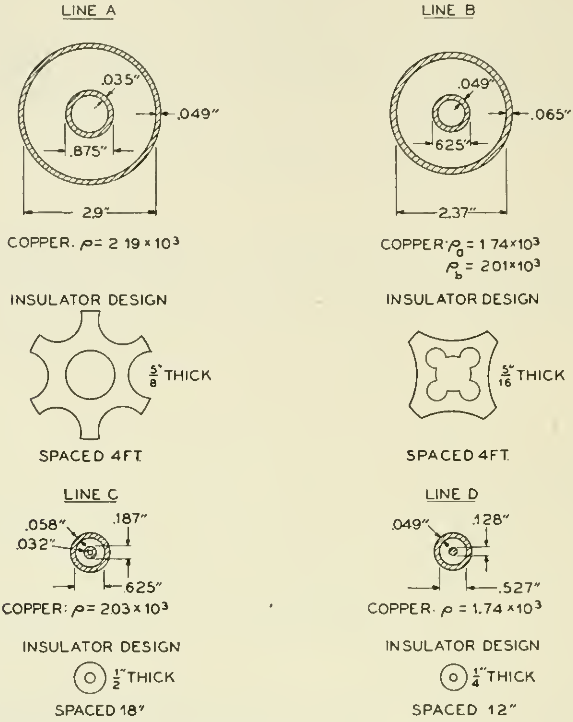


Fig. 1—Principal dimensions of the concentric tube lines upon which the experimental resistance measurements shown in Fig. 2 were made.

Line *E*—Same dimensions as "C." Brass ($\rho = 6.6 \times 10^3$) outer pipe.
Line *F*—Same dimensions as "C." All brass ($\rho = 6.6 \times 10^3$), insulators spaced 22 inches.
Line *G*—Same as "C" but filled with insulators.
Line *H*—Lead sheath cable, No. 18 B & S copper ($\rho = 1.7 \times 10^3$), rubber insulation, lead ($\rho = 17 \times 10^3$) $\frac{1}{8}$-inch inside diameter.
Line *I*—Same as "C" but insulators spaced 9 inches.
Line *J*—Same as "F" but insulators spaced 18 inches.

The results of these measurements appear on Fig. 2. The solid curves were computed by means of (3), neglecting the conductance term. The points are the experimental observations. Note that, excepting lines *G* and *H*, the small margin between measured and calculated values shows that the leakage losses are very small. It is believed that the scattering of the observed points could have been reduced appreciably if corrections for variations of resistivity with temperature had been made.

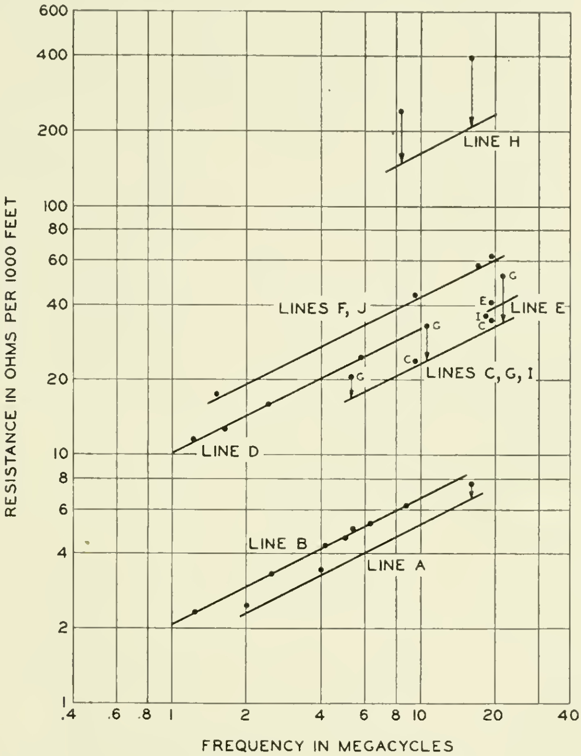


Fig. 2—Radio-frequency resistance measurements upon concentric tube lines. The curves are based upon computations and the points are experimental observations. See Fig. 1 for line details.

In lines *B*, *E*, and *H* the sheath and the inner conductor comprised materials of different resistivities. Calculations for these cases were made with the assistance of a modified form of (3):

$$R = \frac{1}{a}\sqrt{\rho_a f} + \frac{1}{b}\sqrt{\rho_b f}, \tag{3a}$$

in which the subscripts denote the inner and the outer conductors.

Line *G* was completely filled with porcelain insulators and line *H* was a rubber insulated, lead sheathed, cable. The curves of Fig. 3 were derived from the difference between the observed and calculated resistance in these two cases. The results so obtained are a fair approximation of the leakage losses. Note that the curves are nearly proportional to the frequency which is to be expected if for a constant voltage the dielectric absorbs a fixed amount of energy each cycle.

As may be expected the velocity of propagation for both lines *G*

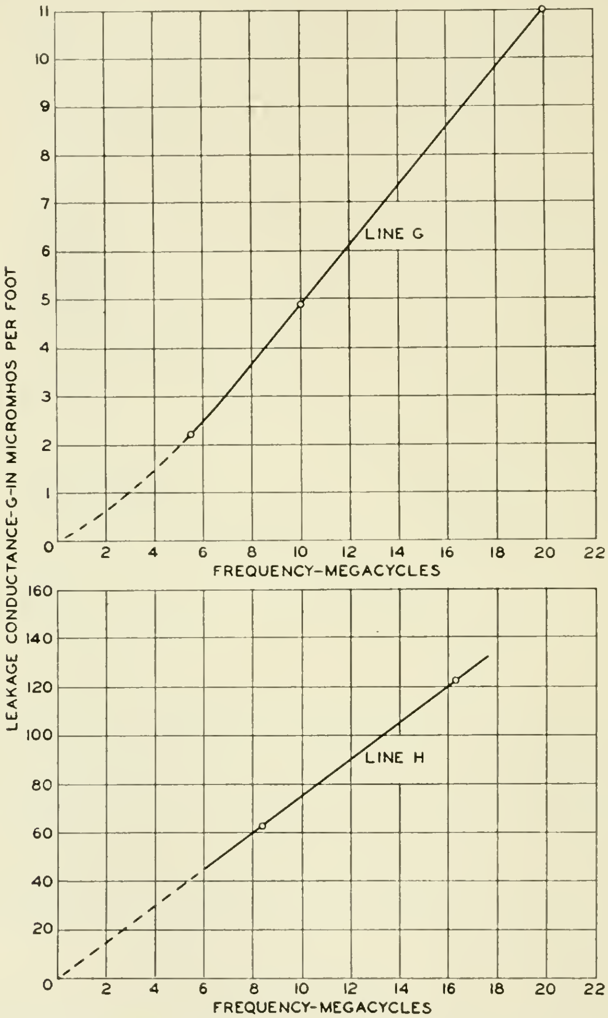


Fig. 3—Derived leakage conductance for the concentric tube lines *G* and *H* of Fig. 1. Line *G* was completely filled with porcelain insulators. Line *H* comprised a No. 18 B & S conductor with rubber insulation and lead sheath.

and *II* was reduced by a factor of approximately 1.8 which corresponds to a dielectric constant of about 3.2. Line *I* which was made with insulators spaced at 9-inch intervals was the only other line which showed a pronounced reduction in the velocity of propagation, the factor in this case being 1.18.

It is of interest to observe that if for economic reasons the diameter of the outer conductor is fixed there is an optimum inner conductor size for minimum attenuation. Employing (1a), (2d), and (3) the real part of the propagation constant may be written as:

$$\alpha = \frac{\sqrt{\rho\mu f}}{276} \frac{\left(\frac{1}{a} + \frac{1}{b}\right)}{\log_{10} \frac{b}{a}} \times 10^{-9}. \tag{4}$$

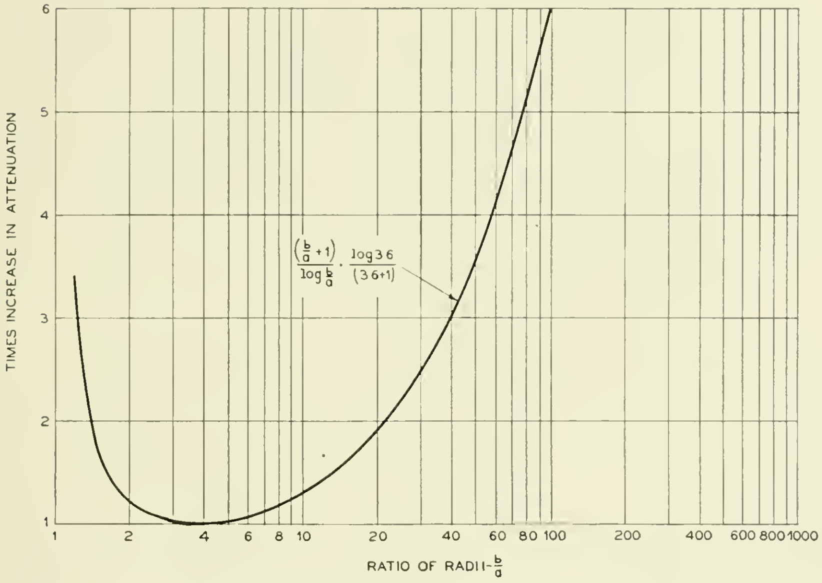


Fig. 4.—The most desirable ratio (36) of outer to inner conductor and the penalty incurred in departing from this value.

This neglects leakage loss and assumes that both conductors are made of the same material. Upon minimizing with respect to a, the optimum ratio:[6]

$$\frac{b}{a} = 3.6 \tag{4a}$$

is readily obtained. This ratio corresponds to a characteristic impedance of 77 ohms. Fig. 4 gives the manner in which the attenuation varies as a function of b/a. Note that a moderate departure from the optimum ratio does not greatly increase the line losses.

---

[6] An experimental figure for the optimum ratio was given by C. S. Franklin in a British Patent, No. 284005. The above derivation for the optimum ratio was disclosed to the writers by E. I. Green and F. A. Leibe, American Telephone and Telegraph Company, New York City.

So far it has been tacitly assumed that the conductors were exactly concentric. Eccentricity affects all of the line constants.[7] However, experience has shown that the departures from concentricity usually encountered in practice produce no appreciable increase in the attenuation constant of the line.

At commercial installations the actual power loss in the terminated line is measured directly in decibels and has invariably been found to agree with the predictions within the precision of such measurements which is about 0.5 db, in field work. Certain of the lines referred to in this paper have been similarly tested in the laboratory under more favorable conditions and yielded agreements within 0.3 db where the total loss was of the order of 4 to 6 db.

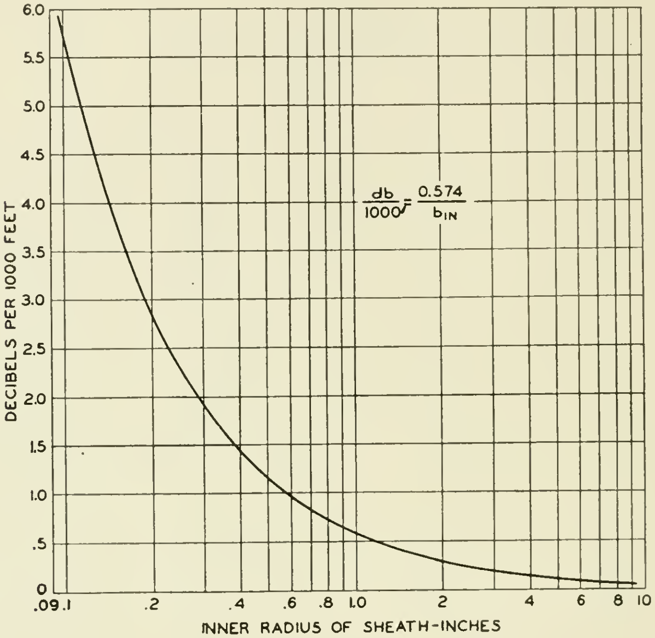


Fig. 5—Calculated losses at 20 megacycles expressed in decibels for concentric-tube lines constructed from copper and employing the optimum ratio (3.6) of outer to inner conductor.

Briefly summarizing it may be said that the attenuation in well constructed concentric lines is proportional to the square root of frequency, inversely proportional to the diameters (optimum ratio) and proportional to the square root of resistivity. Numerically, the loss for copper lines of optimum ratio, neglecting leakage, is:

---

[7] A. Russell, "Alternating Currents," Vol. 1, p. 166, Cambridge Press.

$$\frac{db}{1000 \text{ ft.}} = \frac{0.128\sqrt{f_{\text{mc.}}}}{b_{\text{in.}}}. \tag{5}$$

A plot of (5) for one particular frequency ($f = 20$ mc.) is shown in Fig. 5.

It is important to emphasize one precaution in the use of concentric lines. The high degree of isolation afforded by concentric-tube lines may be easily destroyed. Owing to pick-up from near-by antennas or from other sources, currents of appreciable magnitude may be flowing upon the exterior of the sheath. Spurious couplings between the antenna and the line or between the equipment and the line may introduce these currents into the shielded circuit. In this manner the discrimination of a receiving circuit against undesired signals may be destroyed. Also, the currents flowing upon the exterior of the sheath may destroy the directional characteristic of the antenna to which the line is connected. Grounds placed at frequent intervals are useful in reducing these currents. Sometimes it is both desirable and convenient to bury the line in the earth. Additional improvement is obtained by constructing the circuits which transform the antenna impedance to the line impedance so as to obtain rigorous symmetry to ground.

III. OPEN-WIRE LINES

The losses in open-wire lines may not be determined in as simple a manner or with the degree of certainty that is possible with concentric-tube lines owing to the complex nature of the electromagnetic field about open-wire lines. The high-frequency resistance of one conductor may be obtained from the foregoing equation (3) by assuming that the radius of the outer pipe is infinite. The characteristic impedance of balanced open-wire lines is obtained with sufficient accuracy from:

$$Z_0 = 276 \log_{10} \frac{2D}{d} \text{ ohms}, \tag{2e}$$

in which D is the axial spacing and d is the wire diameter. Some typical results for the resistance of a single conductor appear in Fig. 6.

At first thought it would appear that, owing to the high resistance of a single conductor, the losses in open-wire lines are higher than in concentric-tube lines. In practical constructions, however, open-wire characteristic impedances 5 to 10 times greater than those for concentric-tube lines are easily obtained. For example, the loss of a 77-ohm concentric-tube line is 6.38 times as great as that for a 770-ohm open-wire line in which the wire diameter is equal to that of the inner conductor of the concentric-tube line. Thus, the attenuation constant for a practical open-wire line may be approximately the same as that for the larger practical sizes of concentric-tube lines. Some typical computations appear in Fig. 7. A balanced two-wire line of 600-ohm characteristic impedance was chosen for the computations.

In Fig. 7 it was assumed that the proximity effect, that is, the redistribution of currents owing to the presence of the second conductor, is a correction of negligible magnitude. Only in the case of large conductors closely spaced does the proximity effect perceptibly increase the resistance. This may be seen from Fig. 8 which shows the increase in resistance due to the proximity of the conductors. There are several excellent published articles upon this subject.[8,9,10,11]

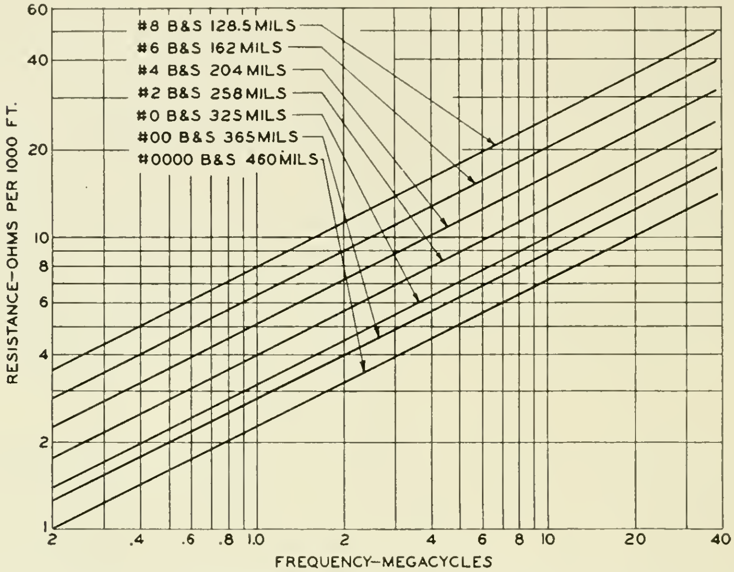


Fig. 6—Calculated radio-frequency resistance for several common sizes of solid copper conductors. Values are for one conductor only.

The foregoing results give, of course, only the power dissipated in copper losses and tell nothing about radiation losses. If the line spacing is less than 1/10 of a wave-length and if the line length is more

[8] J. R. Carson, *Phil. Mag.*, Ser. 6, Vol. 41, p. 607, April, 1921.

[9] H. B. Dwight, *Jour. A. I. E. E.*, p. 203, March, 1922.

[10] H. B. Dwight, *Jour. A. I. E. E.*, p. 827, September, 1923.

[11] S. Pero Meade, *Bell Sys. Tech. Jour.*, Vol. 4, No. 2, April, 1925. The equations given in this reference were employed in computing Fig. 8.

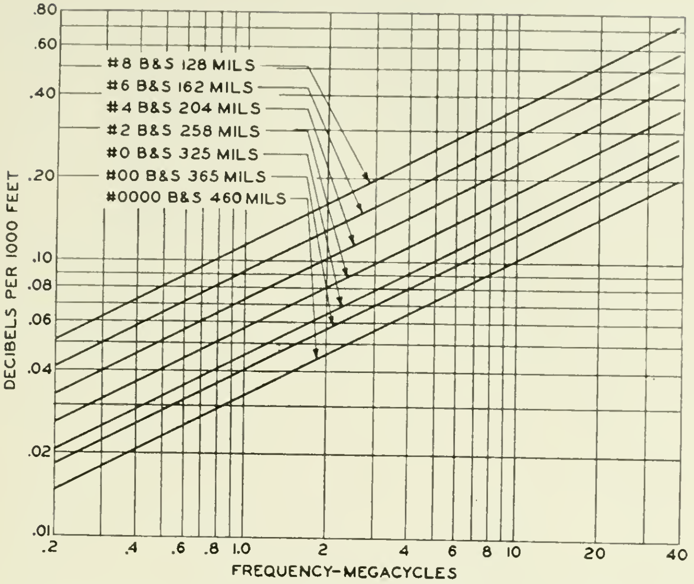


Fig. 7—Calculated attenuation expressed in decibels for copper losses in 600-ohm lines made up from common sizes of solid conductors.

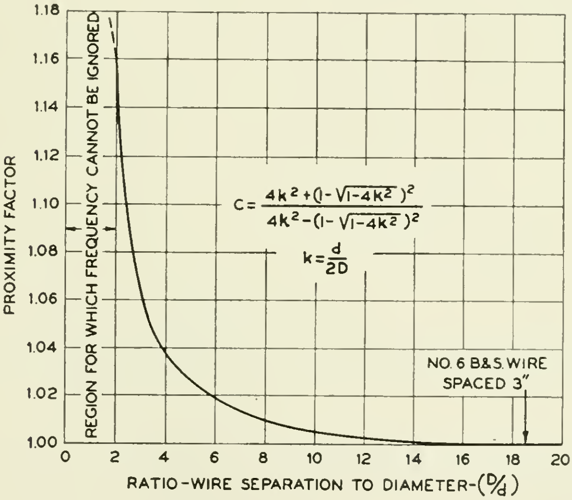


Fig. 8—Calculated values for the increase in copper losses due to the redistribution of current at close conductor spacings in balanced two-wire lines.

than 20 times the line spacing, the power radiated by a two-wire line terminated in its characteristic impedance is approximately:

$$\frac{P}{I^2} = 160 \left[\frac{\pi D}{\lambda}\right]^2 \text{ watts/(amperes)}^2, \tag{6}$$

in which:

D/λ is the line spacing and
I is the r.m.s. value of the current in the line.

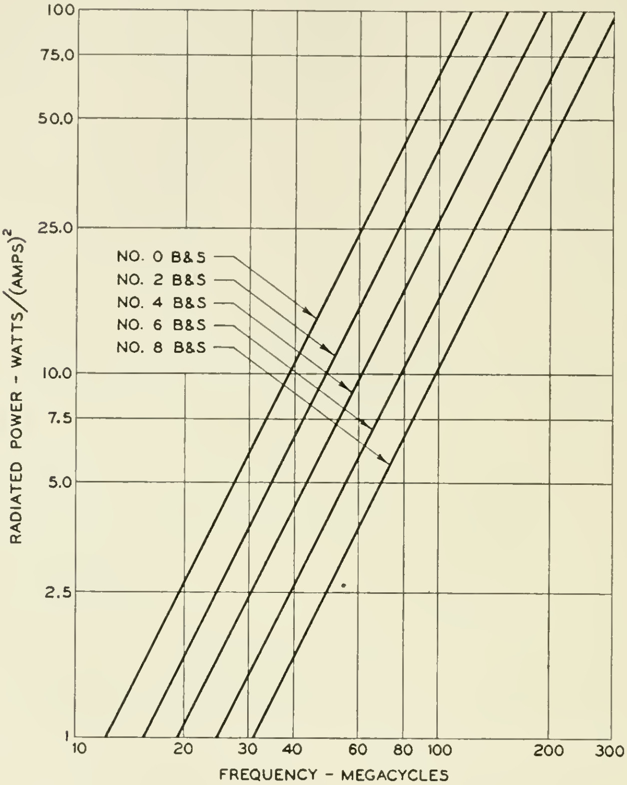


Fig. 9—Calculated power radiated by 600-ohm lines terminated in the characteristic impedance for several common conductor sizes. Note that line length is not involved.

Under these conditions the radiated power is independent of line length. More accurate equations appear in an appendix to this paper. It may be concluded from (6) that the power radiated by a terminated line is, in magnitude, approximately twice that radiated by a doublet antenna of length equal to the line spacing. Thus, the most simple circuit with which it is possible to terminate an open-wire line, a resistance of length equal to the line spacing, will radiate approximately one half as much power as the line. Therefore, considering both load and generator terminations, the total power dissipated in radiation may be approximately twice that given in (6). The equation is plotted on Fig. 9 for the cases of several 600-ohm lines constructed from practical conductor sizes. It may be seen from this figure that the power radiated by a practical terminated line is negligible as compared to the power transmitted by the line provided that operations are confined to wave-lengths other than those in the ultra-short-wave region.

If the currents in the two wires are unequal or are not exactly 180 degrees out of phase there is an appreciable amount of power radiated by a two-wire line. Unbalances of this kind become evident when the driving voltages, measured to neutral, are incorrectly balanced and phased. Such unbalances also arise if the voltages induced by the antenna set up currents in the line which employ the two conductors in parallel.

For the purpose of computation unbalanced currents may be considered as flowing in a single conductor parallel to a perfectly reflecting earth. The amplitude of the current in the single wire may be assumed to be the vector sum of the current values in the two conductors. This procedure ignores the mutual interactions of the balanced and unbalanced currents flowing in the two-wire line and hence, the results so obtained are not strictly correct. It is believed, however, that the error is small.

Based upon these assumptions the power radiated by unbalanced currents is approximately:[12]

$$\begin{aligned}\frac{P}{I^2} = 30\Bigg[\,&0.5772 + \log_\epsilon(2L) - \sin^2(L)\left(1 - \frac{\sin H}{H}\right) - Ci(2L) \\ &- 2Ci(H) + Ci(\sqrt{L^2 + H^2} - L) + Ci(\sqrt{L^2 + H^2} + L)\Bigg],\end{aligned} \tag{7}$$

in which:

P/I^2 is expressed in watts/(amps)^2

I = r.m.s. value of current at a position along the line of maximum current,

$H = \frac{4\pi h}{\lambda}$, $\frac{h}{\lambda}$ being the height of the wires above ground in wave-lengths,

and

$L = \frac{2\pi l}{\lambda}$, $\frac{l}{\lambda}$ being the length of the line in wave-lengths.

---

[12] See appendix.

The equation is plotted on Fig. 10 for two specific heights above ground and for various line lengths. Upon examining Figs. 9 and 10 it may be concluded that in practical constructions a thirty per cent unbalance in line currents radiates an amount of power roughly equal to that radiated by the balanced currents in the line.

It is our experience that losses due to current unbalances are appreciably greater and somewhat different in character from those indicated by (7). The discrepancy may reside in the assumptions employed in deriving the equation. In particular, the losses in the earth have been ignored. It may well be that the soil over which the line is erected introduces large losses in the line, particularly when the currents are unbalanced. Such losses would augment the attenuation constant of the line.[13] At least, the computations indicate the desirability of maintaining careful line current balances.

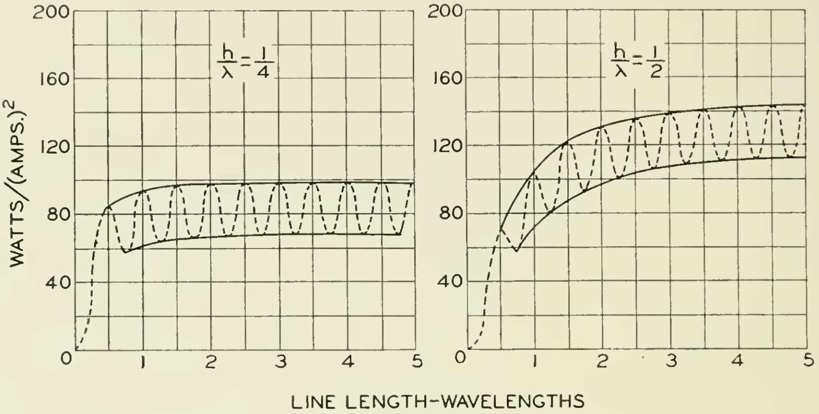


Fig. 10—Approximate power in watts radiated by an unbalanced current of 1.0 r.m.s. amperes in a long two-wire line. Also, the power radiated by a single wire parallel to the earth for 1.0 r.m.s. ampere line current. Two cases, $\frac{1}{2}$ and $\frac{1}{4}$ wavelengths above ground are illustrated.

Some remarks upon the proper procedure for inserting the power losses due to radiation into the equations for the line may be of interest. Carson [14] has shown that the conventional solution of the transmission equation for guided waves on wires is incomplete and does not explain the phenomena of radiation. He shows that a "principal wave," and hence the currents in the conductor associated with this wave, travel along the conductors without sensible attenuation due to radiation. Radiation from the line results in the attenuation of an infinite number of "complementary waves." These are highly attenuated so that the radiation of energy is a phenomenon essentially associated with the terminals of the line or points of discontinuity which set up reflected waves.

---

[13] John R. Carson, *Bell Sys. Tech. Jour.*, Vol. V, No. 4, October, 1926.

[14] John R. Carson, *Jour. A. I. E. E.*, p. 908, October, 1924.

It may be concluded from Carson's mathematical investigation that the radiation resistance is a term to be added to the impedance of the line at the terminals or points of discontinuity and that it does not appear in the propagation constant. On this basis, the power radiated by a practical balanced transmission line is negligibly small when compared to the power being transmitted by the line except perhaps for operation at the very short wave-lengths.

Experimental data for the attenuation in open-wire lines which are as complete as those already shown for concentric-tube lines are not available for this paper. Some typical observations for 600-ohm lines constructed with No. 6 B & S semi-hard drawn copper wire appear in Fig. 11. The points are experimental observations. The lower curve was computed for the case 1830 e.m.u. copper resistivity. The observed values are about 66 per cent higher than the computed values.

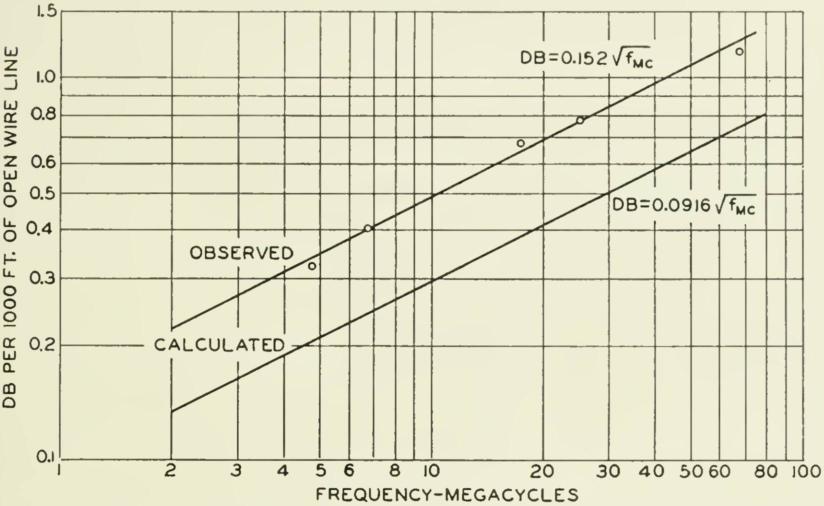


Fig. 11—Experimental observations of attenuation in a 600-ohm line comprising 0.162-inch copper conductors. The points are observed values. The lower curve is calculated only on the basis of copper losses.

The experimental procedure was as follows. A line 2000 feet long was carefully balanced and terminated by an iron wire line [15] for each of the experimental observations. By means of a portable calibrated indicating device the average currents for the one-half wave-length of line at the near end and at the far end were obtained. The attenuation in decibels was computed from average near end and far end current ratios.

---

[15] See Section VII.

It is difficult to explain the discrepancy between observed and computed values. If the resistivity employed in the computations were to be increased from 1830 to 5030 e.m.u. (a multiplication factor of 2.75) the computed curve so obtained would be in good agreement with the observed results. It is true that the wires were somewhat weathered. There is, however, little reason to believe that an appreciable amount of current flows in the oxide layer covering the wires. Effects of this kind would have been evident in the measurements upon concentric-tube lines. It already has been mentioned that small current unbalances in the line may produce losses in the earth which increase the real part of the propagation constant. Possibly, losses of this kind may explain the discrepancy.

IV. Notes on Matching Impedances

It already has been mentioned that standing waves on a transmission line augment line losses. The penalty which is imposed by improper impedance matches may be seen from Fig. 12. This figure plots line loss as a function of the degree of matching for several attenuation factors. The line loss is computed from the ratio of the power dissipated in the load to the total power obtainable from the generator. The curves were obtained from conventional transmission line theory. For the purpose of simplifying calculations the line length is assumed to be an integral number of one-quarter wave-lengths, thereby eliminating complex impedances. Otherwise the length of the line is immaterial, the product of length and attenuation per unit length being the criterion of loss.

In the diagrams of Fig. 12A and Fig. 12B the circuit M is an adjustable ideal transformer. For every value of the resistance R the transformer M is assumed to be adjusted so as to maximize the load power. This process is equivalent to matching impedances at the terminals of the line adjacent to the transformers. It is of interest to observe that where R is not equal to the characteristic impedance Z_0, this adjustment does not yield an impedance match at the line terminals remote from the transformers. It is of further interest to observe that in the case where the load impedance is variable (Fig. 12B) the optimum adjustment is a compromise between a non-reflecting termination and an impedance match at the generator end.

Conventional tuned transformers may be employed to match the line impedance to the antenna and radio equipment impedances. In the transmitting case, tuned circuits are often found to be both bulky and costly. There are a number of schemes which employ a short section of line as a transformer element. These are feasible only at frequencies for which the wave-length is short. The circuits so provided are extremely simple and cheap.

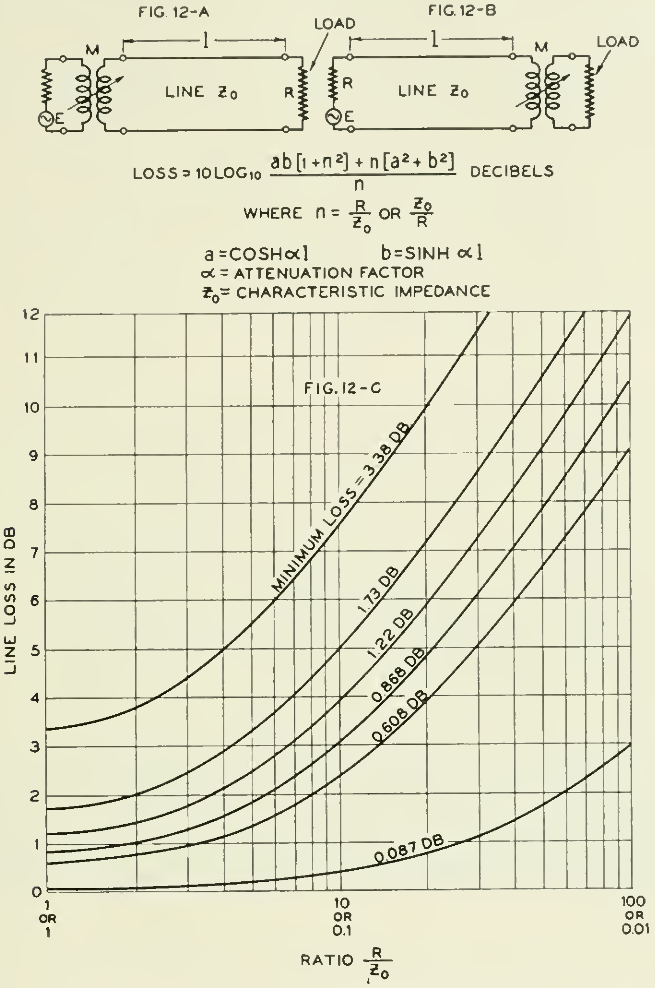


Fig. 12—Line loss as a function of the degree of matching. The several curves are designated in terms of the minimum line loss obtained for the case of a perfect match with the line characteristic impedance.

In another paper [16] a scheme for employing a one-quarter wave-length section of line as a step-up or step-down transformer was described. Briefly the principle of operation is the fact that the sending end impedance Z_s and the receiving end impedance Z_r are related to the characteristic impedance Z_0 by the simple expression:

---

[16] E. J. Sterba, *Proc. I. R. E.*, p. 1184, July, 1931.

$$Z_s Z_r = Z_0^2. \tag{8}$$

Thus, by choosing the proper characteristic impedance any two real impedances may be matched provided these do not differ too greatly.

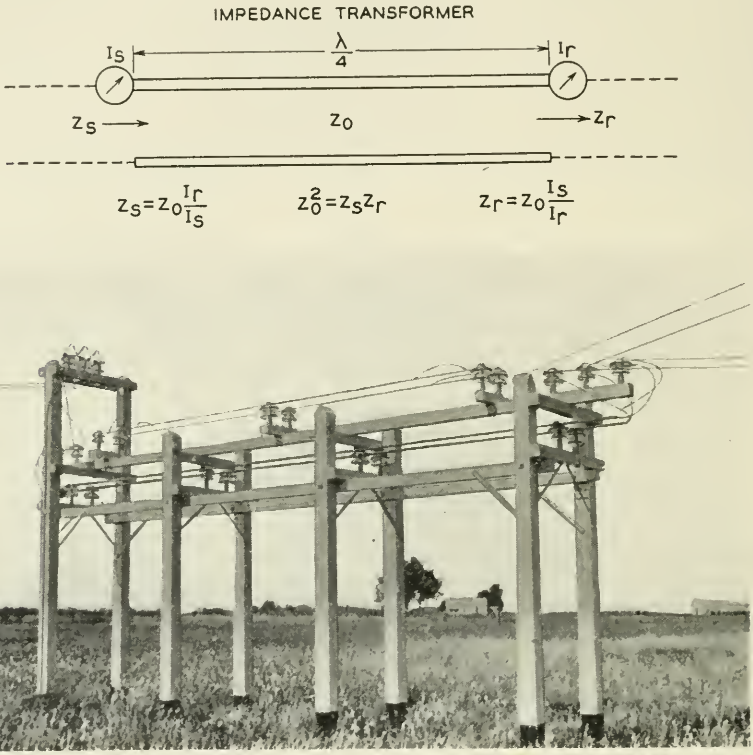


Fig. 13—The principles underlying the use of one quarter wave-length bars as a transformer are shown in the upper diagram. The lower illustration depicts a commercial installation.

Often the scheme is made workable by constricting the line spacing for a one-quarter wave-length section. Where a large difference in transformer line and transmission line spacing is undesirable the transformer line may comprise conductors of large diameter. A transformer set-up of this type is shown in Fig. 13.

There is another effective way for transforming line impedances by means of short line devices.[17] A complex impedance at the proper position along a partially terminated line is selected such that a shunt reactance at this position transforms the real part of the impedance to the surge impedance of the line at essentially unity power factor. The shunt could, of course, be a lumped reactance. It is found convenient to employ a short section of line for this reactance. A position along the line for the shunt reactance of either leading or lagging power factor may be chosen. In the former case the shunt reactance must be inductive and in the latter case capacitive. Computed shunt impedance positions for the two cases and the values of the shunt impedance in terms of line length for 600-ohm lines and various standing wave amplitudes on the unterminated section appear in Fig. 14. Actual settings correspond very well with the calculated settings.

V. RESISTANCE AND ATTENUATION MEASUREMENTS ON TRANSMISSION LINES

The following is a description of some of the measurement methods which have been found useful in the study of transmission lines. The schemes may not be applicable to every phase of the transmission line roblem. However, it is hoped that they may suggest precautions top be observed in performing transmission line studies.

One scheme, very commonly employed, is to measure the attenuation along a transmission line by actual current measurements. This method is particularly suited to measurements upon a long line terminated in its characteristic impedance. It has been found desirable to measure the current amplitudes at close intervals for at least a one-half wave-length section at the near end and the far end of the line. In this manner an average result which reduces observational errors and errors arising from standing waves of small amplitudes is obtained. From the ratio of the average sending end current I_s and the average receiving end current I_r and the average distance l between the two sections of line the attenuation per unit length is obtained from the definition:

$$\text{db} = 20 \log_{10} \frac{I_s}{I_r}, \tag{9}$$

and since

$$\frac{I_s}{I_r} = \epsilon^{Rl/2z_0}, \tag{10}$$

$$\frac{\text{db}}{l} = 4.343 \frac{R}{Z_0}, \tag{11}$$

---

[17] Disclosed to the writers by P. H. Smith, Bell Telephone Laboratories, Inc., New York, N. Y.

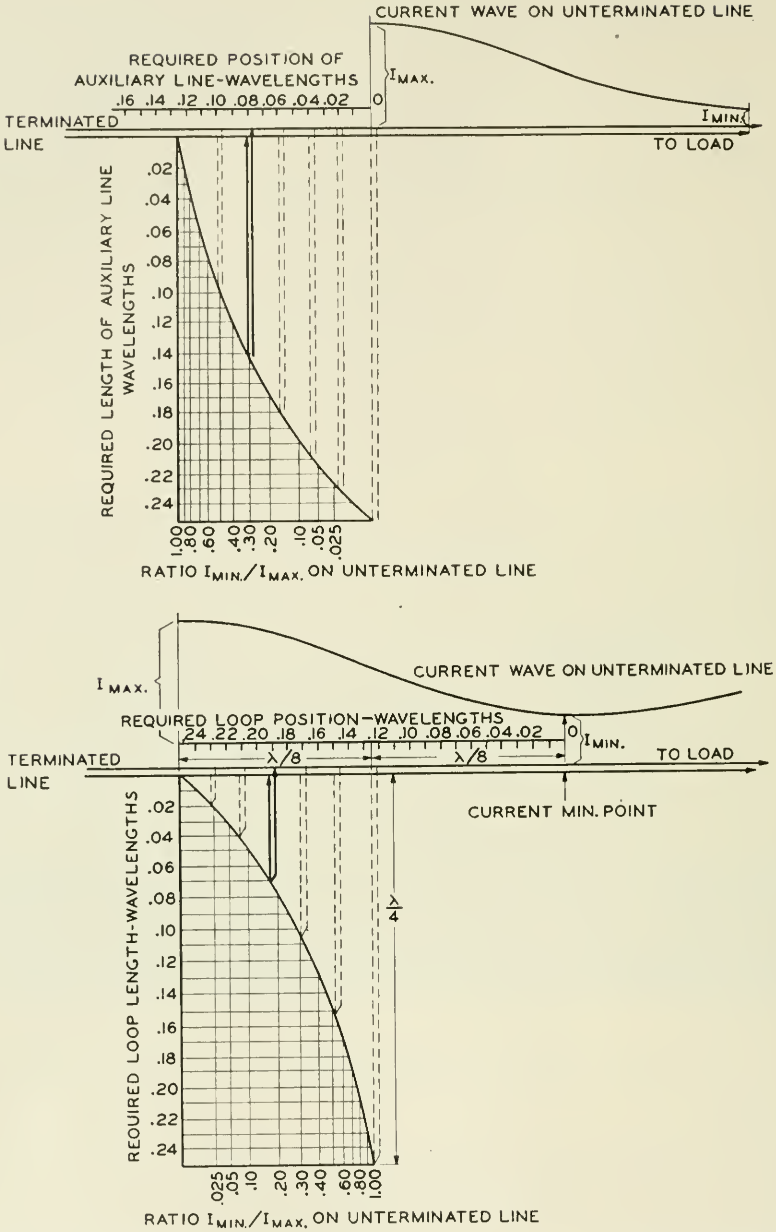


Fig. 14—The use and adjustment of an auxiliary line as a transformer element. The settings are computed for the case of 600-ohm lines. The position and length of the auxiliary line may be obtained from the curves for any given ratio of minimum to maximum currents on the unterminated portion of the line.

from which the resistance R per unit length may be obtained with a degree of accuracy depending chiefly on how accurately the characteristic impedance Z_0 is known.

The current distribution along the line is most conveniently obtained by means of a portable indicating device. Three designs which have been found useful are shown in the following figures. The indicator to the left of Fig. 15 is used for measurements upon open-wire lines. The manner in which it operates is evident from the figure. Note that except for closely spaced wires the device is sensitive only to the current in one side of the line. When this device is used the currents in both sides should be measured to assure that there are no large current unbalances and for the purpose of averaging out any small unbalances. The device in the center of Fig. 15 is coupled to both sides of the line and is not sensitive to currents which employ the two-line conductors

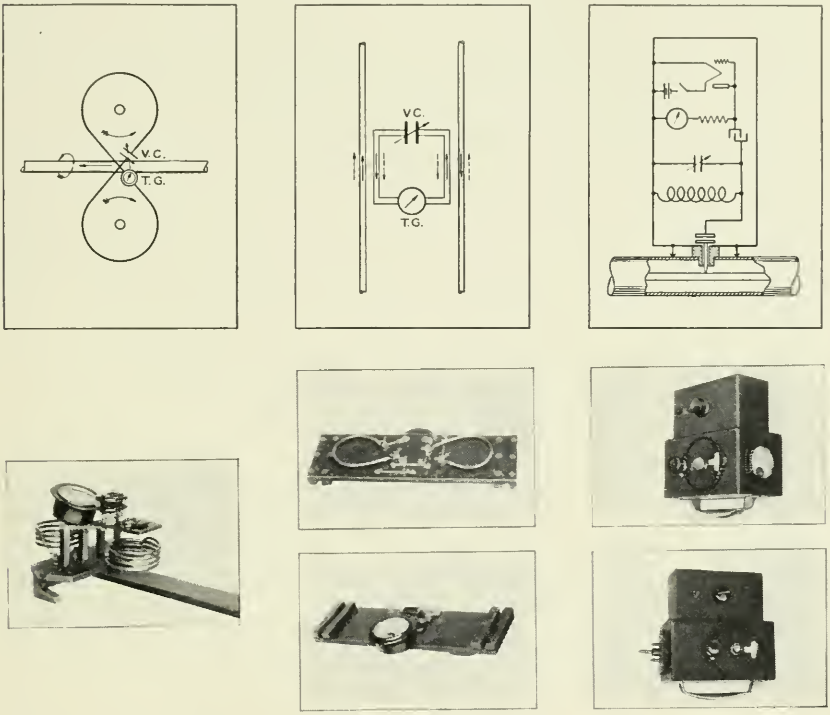


Fig. 15—Three types of portable indicating devices for observing the current distribution on transmission lines. The device to the left is sensitive chiefly to the current flowing in one conductor. The device in the center responds chiefly to balanced currents and is not sensitive to unbalanced currents. The device to the right is suitable for observing the voltage amplitudes on concentric tube lines.

in parallel. It is useful where for other reasons the unbalanced currents cannot be reduced to a desirably low value. The device shown to the right of Fig. 15 is suitable for measurements on concentric lines. In order to employ this device openings at regular intervals are required in the outer sheath. An important precaution to be observed in employing this last device is that the shielding be sufficiently thorough to assure no pick-up from stray currents flowing upon the outside of the sheath.

It is of course essential that all portable devices of this kind extract a very small proportion of the power in the line; otherwise, the device becomes a source of reflection and spurious results are obtained.

Another method of measuring the attenuation of a line which is particularly useful in studying the effects of current unbalances is to employ a small portable horizontal antenna the impedance of which matches the characteristic impedance of the line. The antenna is connected in a short section and then in a long section of the line. It is essential that the height of the antenna above ground be equal for the two positions. Also, the location for the experiment should be such that the same ground losses are present for the two positions. The ratio of the antenna currents for the two positions and for the condition of equal power input is a measure of the total line losses.

One of the most satisfactory schemes for measuring line attenuation is the direct measurement of the line sending end impedance by means of the familiar resistance substitution method. It has been used extensively in measurements of concentric lines. For this purpose it is necessary to employ lines either open- or short-circuited at the far end and to restrict the measurements to lines which contain an integral number of quarter wave-lengths.

Conventional transmission line theory indicates that under these conditions the impedance is either:

$$Z_1 = Z_0 \tanh\left(\alpha \frac{n\lambda}{4}\right), \tag{12}$$

or:

$$Z_2 = Z_0 \coth\left(\alpha \frac{n\lambda}{4}\right), \tag{13}$$

where:

Z_0 = characteristic impedance,
α = attenuation factor; i.e., the real part of the propagation constant,
λ = wave-length, and
n = an integer denoting the number of quarter wave-lengths.

If n is even and the termination is a short circuit or if n is odd and the termination is an open circuit, (12) is employed. If n is odd and the termination is a short circuit or if n is even and the termination is an open circuit, (13) is employed. The attenuation factor is given by:

$$\alpha = \frac{1}{2}\frac{R}{Z_0} + \frac{1}{2} GZ_0 \text{ napiers per foot,} \tag{14}$$

$$= 4.34 \left(\frac{R}{Z_0} + GZ_0 \right) \text{ decibels per foot,} \tag{14a}$$

where:

R = resistance in ohms per foot,
G = conductance in ohms per foot, and
Z_0 = characteristic impedance in ohms.

For all the lines concerned with here tanh $[\alpha(n\lambda/4)]$ may be replaced by $[\alpha(n\lambda/4)]$ without more than 1.5 per cent error. Thus, (12) and (13) reduce to

$$Z_1 = \frac{R}{2}\frac{n\lambda}{4}\left(1 + \frac{GZ_0^2}{R}\right), \tag{12a}$$

[18]
$$Z_2 = \frac{2Z_0^2}{\frac{n\lambda}{4}R}\left(1 - \frac{GZ_0^2}{R}\right). \tag{13a}$$

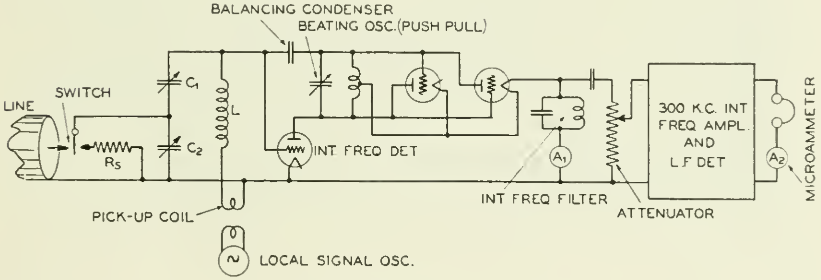


Fig. 16—Schematic diagram of apparatus for measuring line loss by means of a resistance substitution method.

Therefore, except for the small contribution of shunt conductance, Z_1 is independent of Z_0. In cases where G is negligible, the measurement of Z_1 gives directly the high-frequency resistance.

The method of measurement is shown schematically in Fig. 16 and an experimental set-up appears in Fig. 17. The modified high-frequency field intensity measuring unit[19] is a convenient indicating device and source of signal. The intermediate-frequency amplifier with its adjustable gain is very useful in maintaining a desirable level in the last detector which is the indicator. Returning to Fig. 16, the local signal oscillator is the source of voltage actuating the tuned circuit LC_1C_2 in which the substitutions are made. Loose coupling is desirable between the pick-up coil and oscillator. In measuring Z_1, which is a low impedance, the condenser C_2 is set at minimum capacity thus effectively making the comparison in a series circuit. In measuring Z_2, which is of the order of hundreds of ohms, C_2 is used to transform Z_2 into an appropriate series value consistent with selectivity and sensitivity.

---

[18] It is assumed here that the conductance term in (14) is small compared with the resistance term.

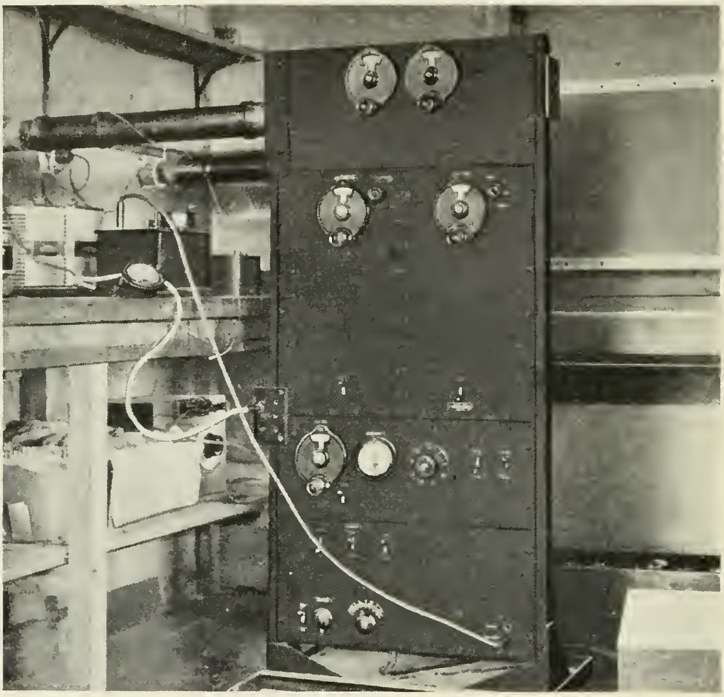


Fig. 17—Experimental set-up of apparatus for measuring line loss.

The comparison resistances used for measuring (Z_1) may be fixed units and the line impedance obtained by interpolation. These comprise straight lengths of high-resistance wire and range from a few tenths of an ohm to ten ohms. The wire size is chosen so that skin effect is negligible. Although these resistances possess appreciable inductive reactance at the higher frequencies the reactance usually may be safely tuned out.

---

[19] Readers not familiar with this measuring unit may refer to a paper by Friis and Bruce, *Proc. I. R. E.*, August, 1926.

It has been found practicable to employ continuously variable resistances made from hard drawing pencil lead equipped with spring clip contacts for measurements of high resistances such as (Z_2). The high resistivity of graphite makes it possible to obtain several thousand ohms in 5 or 6 inches, free of skin effect and with but little inductance.

The foregoing resistance substitution method has been found satisfactory for the purpose of measuring the characteristic impedance of lines. Two methods have been employed. One of these is the familiar procedure in which the sending end impedance is measured for the case of the line open and short circuited at the far end. The geometric mean of the two impedances so obtained is the characteristic impedance of the line.

Another scheme producing more precise results is also adapted to the foregoing resistance measuring method in that it requires that the line be some odd number of one-quarter wave-lengths long. Such a line transforms to a different value a terminating impedance which is other than the characteristic impedance. By comparing a variable terminating resistance directly with the value to which it is transformed by the line a setting may be found for which the line functions as a one-to-one transformer. For this condition the value of the variable terminating resistance is the characteristic impedance of the line. Here again drawing pencil leads have been found to be satisfactory termination resistances when set by direct-current measurement methods.

In practice the foregoing resistance substitution method brings to light many slight irregularities. Variations of apparent characteristic impedance with frequency as much as 5 to 10 per cent have been found for concentric-tube lines equipped with elbows, couplings, and similar fittings. It is believed that impedance variations of this order are to be expected from some such irregularities unless particular care is taken in the construction of the fittings. On the other hand, it has been found that a short straight length of carefully constructed concentric-tube line is so smooth that its characteristic impedance may be employed as a calculable standard.

VI. Practical Construction Details

Open-wire radio-frequency line construction is not very different from that employed in power practice. One outstanding difference is that line supports and insulators must be considered as individual irregularities spaced at intervals often greater than one wave-length. The effect of one such irregularity may be small. The total effect in a long line, however, is sometimes appreciable.

The body of the insulator, since it has a dielectric constant appreciably different from air and since its dimensions are comparable with the line spacing, is in itself a line irregularity. Tie wires or conductor clamps augment this effect. Cross arms and pins employed for mounting pin-type insulators also add to the effect, particularly during wet weather.

From the standpoint of line irregularities suspension-type insulators are more desirable than pin-type insulators. The latter construction, however, appears to be more practical because the lines are more rigid, sway less during wind storms, and because no intermediate spreaders are required to maintain the desired line spacing.

One other difficulty with open wire lines is the drift in velocity of propagation and surge impedance during rain and sleet storms.[16] Since a similar effect occurs in the elements of the antenna there is a decided drop in the efficiency of the combined antenna and line during rain and sleet storms. The effects of sleet may be reduced by heating the wires with sleet melting currents. The conductor size may be increased to reduce the effects of wet weather but this makes sleet melting more difficult.

There is an appreciable pick-up between balanced open wire lines on common supports. It appears desirable to separate lines to a common transmitter by at least 10 times the conductor spacing. Spacings greater than this may be required if two lines are to be operated simultaneously and in some cases it is more desirable to employ separate line supports in order to reduce the possibility of cross-talk difficulties. Of course any current unbalances in two parallel lines greatly increase the danger of cross-talk.

Concentric tube line construction is not as simple as open wire construction. Considering the transmitting case, there is a smaller safety factor for voltage overloads. Insulators are required to withstand high voltage gradients. Temperature changes with ensuing line expansions and contractions must be given consideration. It is these factors in addition to the added expenditure for copper which make concentric line construction more costly than open wire construction.

The first consideration in the design of a concentric line is the weight of the outer sheath. If the line is to be employed for high power transmitting purposes the voltage safety factor may be so low that accidental dents in the sheath may lead to breakdown. Obviously, there is a choice between a large diameter, lightweight sheath and more rugged small diameter sheath without an appreciable difference in copper expenditure. Other factors which involve the remainder of the radio plant often determine the size of the outer sheath. We have found that for outer sheaths a diameter of 2.5 inches and a radial thickness of 0.0875 to 0.10 inch provides lines which are sufficiently rugged for transmitting 15 kw of modulated power at 16 meters wave-length.

Careful consideration needs to be given to the problem of protecting concentric lines from voltage overloads which may be brought about by accidental open or short circuits or by flashovers. Voltages of the order of 30,000 to 90,000 volts may easily be built up in this manner at the shorter wave-lengths. Horn gaps are useful if located in the proper way. It is fortunate that conventional line input circuits are apt to be detuned in the event of an accidental open or short circuit on the line and that very little power may then be transmitted to the line.

Beads of high grade porcelain in diameters up to one inch are satisfactory insulators for low power and receiving lines. However, such simple insulators are not suitable for high power work. Owing to the volume of dielectric in large annular insulators sufficient heating may occur at the higher voltages to destroy the insulator. Insulators such as those described for line *B* Fig. 1 have been found suitable at the higher voltages.

The air film between the insulator and the inner conductor lies in a region of steep voltage gradient. Even under what is considered normal operating voltage there may be enough corona in this region to produce heating of the insulator. It may be of interest to mention that a line approximately as described in *B* Fig. 1 has been found satisfactory for normal operation at 16 meters for a carrier power of 15 kw. The line breaks down in the region of the insulator at 9000 r.m.s. volts.

For transmitting purposes it has been found desirable to employ glazed insulators in concentric tube lines because dirt, soldering fluxes, etc. acquired in assembly operations are more readily removed from glazed insulators.

There are a number of simple ways in which insulators may be held in place in concentric tube lines. For low power work and receiving purposes wire clips, rivets or even extruded metal ears upon the inner conductor, are satisfactory. As a rule these do not prove satisfactory at higher powers owing to high potential gradients at points and sharp edges. Small rings riveted or soldered upon either side of the insulator have proven satisfactory. Lines with soldered rings are more easily repaired. Care must be exercised, however, to prevent condensation of metal and fluxes in the pores of the insulator.

In open wire construction it is customary to accommodate line variations brought about by temperature changes by adjusting the sag of the conductors. Provisions for temperature variations in concentric tube lines are not so simple. The first obvious remedy is to employ lines buried at sufficient depth so that temperature changes are reduced to a slow seasonal variation. At the present time a buried 3/8-inch line has been in service for more than one year without developing faults. Without longer experience with concentric lines we would question the advisability of burying larger lines which are to be employed at high voltages due to the difficulty of finding faults should these occur.

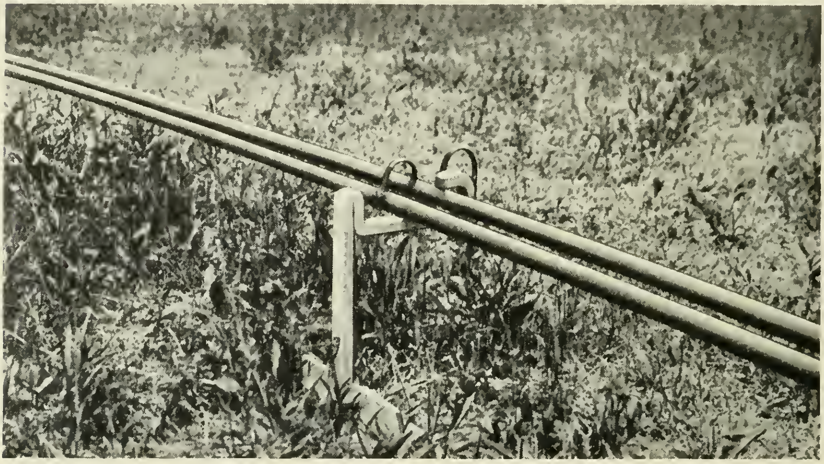


Fig. 18—A short section of three-quarter inch diameter line showing support for holding line in a sinuous form.

A very simple scheme, suitable for small lines, is to reduce the effects of temperature variations by laying the line in a sinuous path as shown in Fig. 18. This construction permits the line to buckle slightly at the curves as the length varies and cumulative changes in length do not appear at the line terminals. The inner conductor is held loosely within the sheath so that it may buckle independently of the sheath. The outer conductor changes its length both at a different rate and at a different time from the inner conductor. With increasing temperature the sheath is at a higher temperature than the inner conductor. There is an appreciable time lag in heating of the inner conductor due to the heat insulation of the air space between the conductors. Small lines laid in a sinuous manner have been found remarkably free from mechanical breakdowns brought about by temperature variations of length.

Sliding joints may be employed to accommodate variations in line length brought about by temperature changes. It is very difficult to make such joints water-tight without recourse to expensive fittings. There is also the possibility of microphonic contacts which are particularly objectionable in receiving work.

The expansion joints shown to the right of Fig. 19 have been employed with some success. Dimensions and shapes should be chosen to minimize the irregularities in line impedance caused by expansion joints. It is a step in the right direction to maintain constant the ratio of conductor diameters at the joint. Even then, it has been found that the irregularities caused by 10 such joints in a 600-foot line are observable (approximately 10 per cent standing waves).

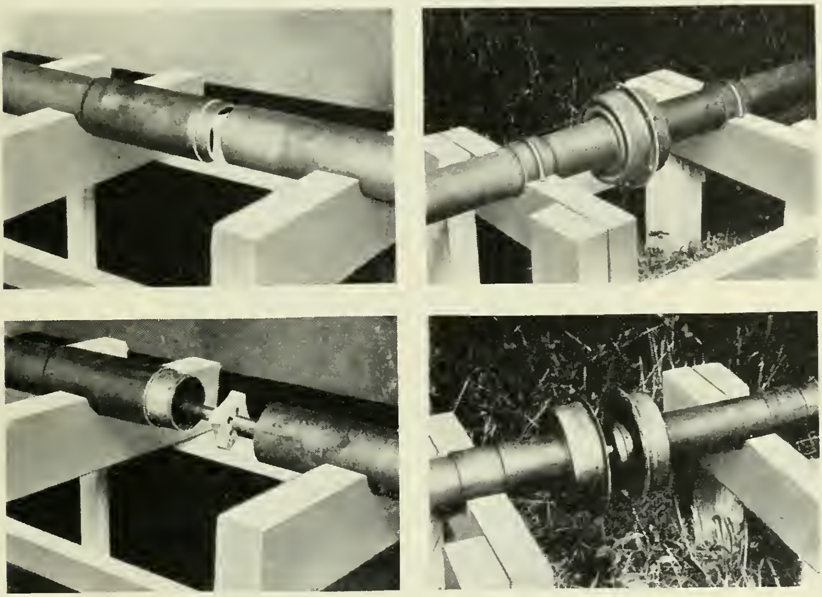


Fig. 19—Experimental expansion and lock joints for large sizes of concentric tube line.

It is necessary that expansion joints be employed in conjunction with lock joints so arranged that no one joint is required to take more than a predetermined portion of the line expansion. One lock joint with an expansion joint 25 feet in either direction has been found to be a satisfactory length within which line variations are corrected. The lock joint proper, Fig. 19, comprises an insulator of the same design as the intermediate insulators but made with an outer diameter equal to that of the outer sheath. It is held in place by a sleeve sweated to the sheath, the sleeve continuing the electric circuit. The insulator is also fixed to the inner conductor by means of rings. Since the lock joint is in a position symmetrical with respect to the two expansion joints it is required to withstand a shearing load brought about only by inequalities in the expansion of the conductors. In order to distribute the line expansion uniformly among the expansion joints it is necessary to clamp the outer sheath of the lock joint to a substantially braced support.

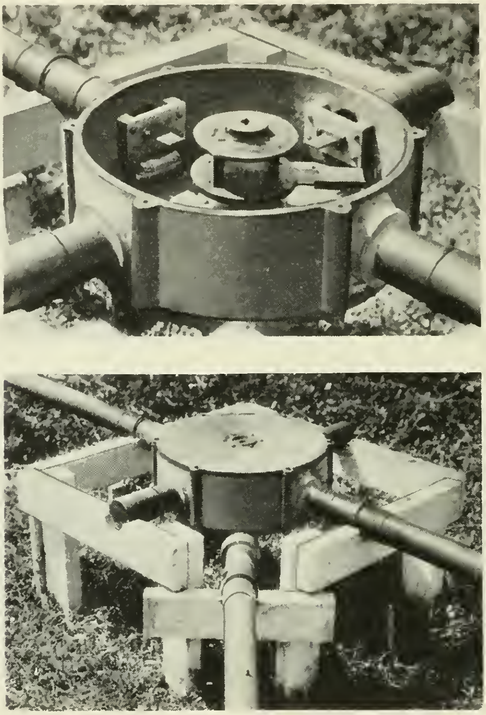


Fig. 20—An experimental selector switch for connecting several antenna lines to one transmitter line. The small coil antiresonates the capacity of the switch for the operating frequency associated with the particular contact to which it is connected.

The joints described above are required to accommodate a total annual variation of 0.5 inches. After one year's period of experimental operation the few faults found in a line containing these joints were nearly all traced to faulty construction at the braced support which clamps the lock joint.

Copper pipe lines may be too costly to permit the installation of more than one line per transmitter. In such cases a selector switch is required if several antennas are to be associated with one transmitter. Some of the details of such a switch may be obtained from the experimental arrangement shown in Fig.20.

The switch is an irregularity on the line and a source of undesired reflections. This difficulty may be corrected by making the design such that capacitive reactance of the switch predominates and then antiresonating this reactance with a suitable inductance. This scheme is effective provided the irregularity is not too great. In the latter event the corrective coil transforms the load impedance to a value different from the surge impedance so that the reflections arising from the mismatch are more serious than from the switch alone.

VII. Other Applications of Transmission Lines

In this section are described a number of transmission line applications to radio work some of which are feasible only at high frequencies because the wave-length is short.

Small concentric lines approximately 3/8-inch in diameter may be employed as radio-frequency wiring in radio stations. Such lines owing to the flexibility of the tubing may be snaked behind partitions in very much the same manner that armored or leaded conductors are installed. For this purpose refrigerator tubing has been found desirable because it is flexible and because it may be procured in long lengths. The inner conductor is insulated from the sheath by means of small porcelain beads spaced at intervals of approximately one inch. The beads are held in position by small metal ears extruded from the inner conductor. The beads fit loosely in the inner conductor so that the line may be bent into arcs as small as six inches radius. Construction details for small concentric lines may be obtained from Fig. 21.

Lines constructed from refrigerator tubing may be buried in the ground. Since only a few splices are necessary the possibility of faults arising from water seeping into the line are correspondingly small. A buried line constructed in this manner has been in service for more than a year without developing faults.

A number of the above-described lines may be terminated upon a jack board and circuits set up with patch cords as in telephone practice. Of course, the beads in the patch cords may be more closely spaced to assure flexibility and freedom from short circuits. The

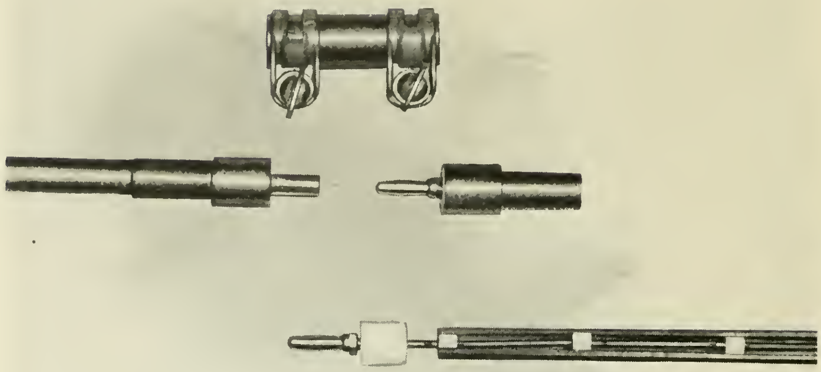


Fig. 21—Details of construction for small concentric lines suitable for station wiring and for patch cords. The plug-and-jack union is an effective scheme for temporarily connecting two small lines.

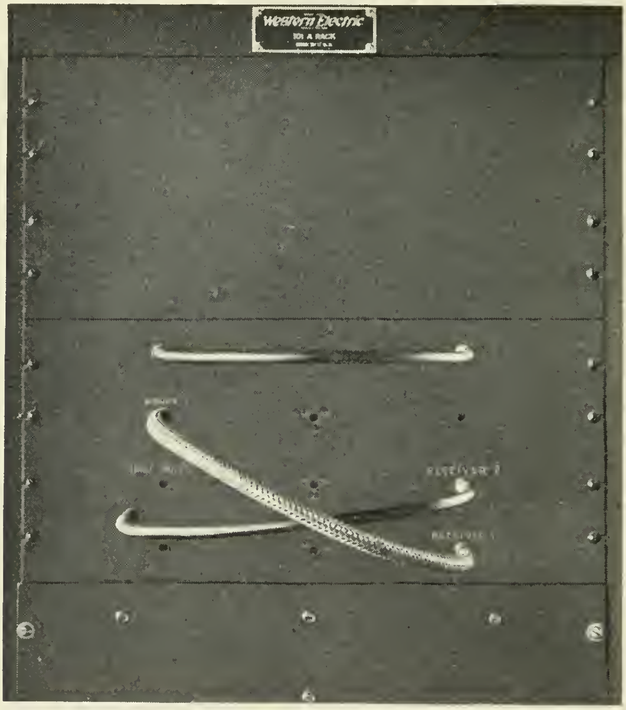


Fig. 22—An experimental radio-frequency jack board terminal for small concentric-tube lines. Path cords are constructed in the manner depicted in Fig. 21.

scheme is particularly advantageous where for operating reasons it is useful to connect any station antenna to a particular receiving unit. A board set up for this purpose is shown in Fig. 22.

Concentric-tube lines may be employed as standards of resistance when other standards become questionable. Since the agreement between theoretical and experimental values of radio-frequency resistance has been found very good at frequencies at high as 20 megacycles the theory may be considered adequate for much higher frequencies. One scheme for utilizing this situation so as to obtain an adjustable

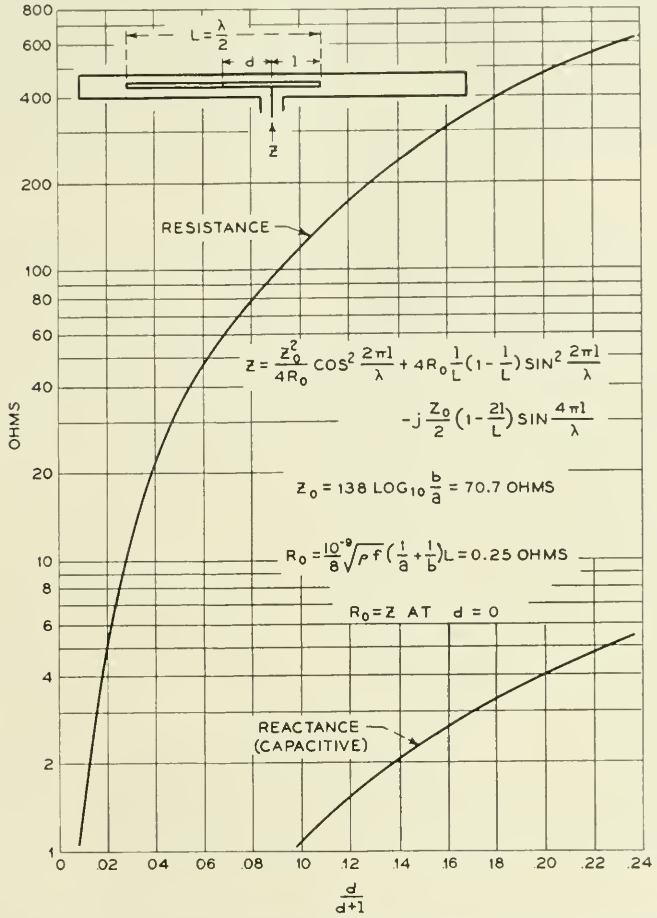


Fig. 23—A scheme for obtaining a calculable radio-frequency resistance standard which is essentially nonreactive and which is adjustable over wide limits of resistance.

radio-frequency resistance will be described. The scheme utilizes the resonant properties of a section of concentric-tube line of which the inner conductor is one-half wave-length long. The required resistance is obtained by a connection to the proper position on the inner conductor. The device is illustrated schematically on Fig. 23. It may be seen from the curves on this figure that the device is a means for obtaining a variable resistance which for most practical purposes is non-reactive. Additional advantages are that the device is rugged and that it may be designed to dissipate an appreciable amount of power.

For the purpose of testing transmitters and for other purposes in which the terminating network is required to dissipate several kilowatts, an iron wire line has been found to be of considerable utility. An iron wire line has the advantage that its impedance is almost independent of the frequency provided that the length of the line is sufficient. This impedance is very closely the characteristic impedance of the line. The far end of the line may be left either opened or closed. For operation at some one frequency the input impedance may be made more nearly equal to the characteristic impedance by means of a termination at the far end. For this purpose the scheme employing a short length of line as a parallel transforming impedance has been found very convenient.

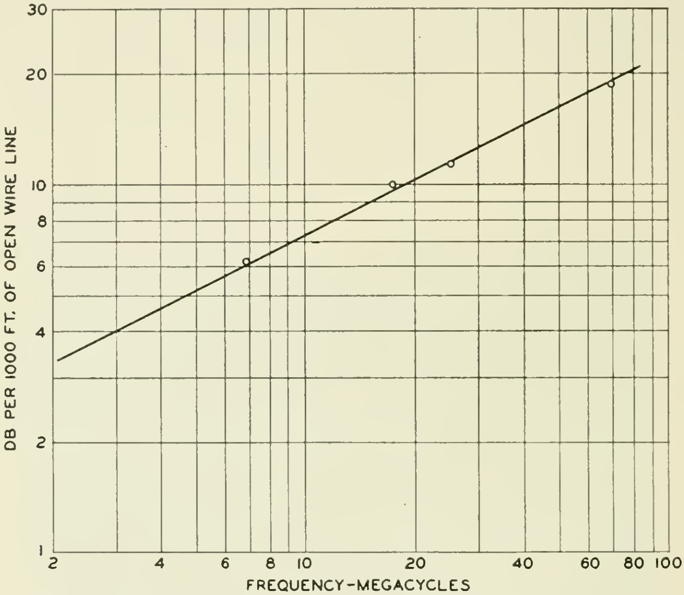


Fig. 24—The curve gives the attenuation in decibels for a balanced 600-ohm line constructed from No. 6 B & S iron wire.

An experimental attenuation curve for a 600-ohm line comprising 0.162-inch iron wire conductors is given in Fig. 24. The current entering the line was approximately one ampere. The measured resistivity for the iron is 12,300 e.m.u. The attenuation could be explained upon the basis that the permeability is 92.[20] This is not an unreasonable value. In fact it may be very desirable to obtain the permeability of iron at radio frequencies by forming the material into a transmission line and observing the line attenuation and direct-current resistance.

An iron wire line 1600 feet long has been in use at the Deal Laboratories for several years for the purpose of testing transmitters. This line successfully dissipates 15 kw. at 20 megacycles.

VIII. Conclusion

In conclusion a few remarks on the relative utility of open-wire and concentric-tube lines may assist in selecting the most desirable construction for a particular service. A definite discrimination between the two is not readily made because the economics of the entire radio plant are involved.

Concentric lines are more costly than open-wire lines. On the other hand, concentric lines permit the installation of a number of radio units within a single structure without incurring difficulties from crosstalk. The first cost and annual charges upon a compact installation may more than offset the cost of the lines when compared with an installation comprising several widely separated structures. Also, concentric lines may be constructed so as to be weatherproof.

There is little choice between the losses in open and concentric lines provided that a reasonable degree of current balance in the open-wire lines is maintained. In order to obtain balances the open-wire line terminal equipment both at the antenna and at the radio unit ends of the line must be carefully designed. The chief source of current unbalance difficulties resides in couplings between the antenna and an open-wire line. These may be materially reduced but cannot be completely eliminated. Another source may be unbalances with respect to neutral at the radio unit.

Complete isolation of the antenna from the line can only be obtained with shielded lines. Similarly, complete isolation from static and other noise sources for which discrimination by the antenna is obtained can only be effected by shielded lines. This is particularly important in reception. In this case small concentric-tube lines with losses as much as 2 db per 1,000 feet may be used provided that the noise level is reduced by a corresponding amount.

---

[20] P. P. Cioffi, Bell Telephone Laboratories, New York City, found an initial permeability of 95 for a sample of the above wire.

We wish to acknowledge the helpful suggestions which have been received in the course of this work from Messrs. H. T. Friis, J. C. Schelleng, and M. E. Strieby. Valuable advice on some of the mathematical questions encountered has been received from Mr. T. C. Fry.

APPENDIX

The following formulas for the power radiated by transmission lines were obtained by the conventional method of postulating the current distribution, calculating the electromagnetic fields and from the fields, the associated radiation by means of Poynting's theorem. As an independent check the same current distribution was postulated and the radiated power calculated following the methods of Pistolkors[21] and Bechman.[22]

Case I

An approximation for a line terminated in its characteristic impedance is a balanced two-wire line carrying a non-attenuated traveling wave. For this case the power radiated is:

$$\begin{aligned} P_1 = 120I^2 \Bigg[& \log_{\epsilon}(2L) - Ci(2L) + \frac{\sin(2L)}{(2L)} + 0.5772 - 1 \\ & - 2Ci(A) + \frac{\sin A}{A} - \frac{\sin(\sqrt{L^2 + A^2} - L) + \sin(\sqrt{L^2 + A^2} + L)}{2\sqrt{L^2 + A^2}} \\ & + Ci(\sqrt{L^2 + A^2} - L) + Ci(\sqrt{L^2 + A^2} + L) \Bigg] \text{ watts}, \end{aligned} \tag{1}$$

in which

$$A = \frac{2\pi a}{\lambda},$$

$$L = \frac{2\pi l}{\lambda},$$

$\frac{a}{\lambda}$ = line spacing in wave-lengths,

$\frac{l}{\lambda}$ = line length in wave-lengths,

I = r.m.s. value of current in each wire, and

$Ci(\ )$ = cosine integral.[23]

The equation simplifies considerably if it is assumed that a/λ is small so that:

$$\sin A \approx A \text{ and } L \gg A,$$

---

[21] A. A. Pistolkors, *Proc. I. R. E.*, p. 562, March, 1929.

[22] R. Bechman, *Proc. I. R. E.*, p. 461, March, 1931.

[23] *See* Jahnke und Emde, "Funktionentafeln."

under which condition:

$$P_1 = 160I^2\left(\frac{\pi a}{\lambda}\right)^2 \text{ watts.} \tag{1a}$$

The numerical constant in (1a) differs somewhat from a result published some time ago by Carson.[24]

Case II

An approximation for an unterminated line is a balanced two-wire line bearing standing waves of the form:

$$I_x = I\cos\left[\frac{2\pi x}{\lambda} + \frac{2\pi m}{\lambda}\right].$$

For this case the power radiated is:

$$\begin{aligned} P_2 = 60I^2\Bigg[& \log_\epsilon(2L) - Ci(2L) + 2\cos M\cos(L-M)\frac{\sin L}{L} + 0.5772 \\ & - 2Ci(A) + \frac{\sin A}{A}[\cos^2 M + \cos^2(L-M)] \\ & - \cos^2 M - \cos^2(L-M) - 2\cos M\cos(L-\text{M})\frac{\sin\sqrt{L^2+A^2}}{\sqrt{L^2+A^2}} \\ & + Ci(\sqrt{L^2+A^2}-L) + Ci(\sqrt{L^2+A^2}+L)\Bigg] \text{ watts,} \end{aligned} \tag{2}$$

in which $M = 2\pi m/\lambda$.

If as before it is assumed that the spacing is small and the line long the equation reduces to the following cases:

Case II-A

When the current is zero at both ends of the line, then,

$$\sin L = 0 \quad \text{and} \quad \sin M = \pm 1$$

and the radiated power is:

$$P_2 = 120I^2\left(\frac{\pi a}{\lambda}\right)^2 \text{ watts.} \tag{2a}$$

This agrees with a result published by Manneback.[25]

---

[24] John R. Carson, *Jour. A. I. E. E.*, p. 789, October, 1921.

[25] Charles Manneback, *Jour. A. I. E. E.*, p. 95, February, 1923.

Case II-B

When the current is zero at one end and maximum at the other end of the line, then,

$$\sin M = \pm 1 \text{ and } \cos L = 0$$
$$\text{or } \cos M = 1 \text{ and } \sin L = \pm 1$$

and the radiated power is:

$$P_2 = 80I^2\left(\frac{\pi a}{\lambda}\right)^2 \text{ watts.} \tag{2b}$$

Case II-C

When the current is maximum at each end of the line, then,

$$\sin M = 0 \quad \text{and} \quad \cos L = \pm 1$$

and the radiated power is:

$$P_2 = 40I^2\left(\frac{\pi a}{\lambda}\right)^2 \text{ watts.} \tag{2c}$$

The approximation for the power radiated by unbalanced currents is essentially the case of a long wire parallel to a perfect earth. The approximation may be obtained from (2) by assuming that power is radiated only in one hemisphere, which divides the numerical constant by a factor of two and by writing for a the quantity $2h$, h being the height of the wire above ground. Equation (7) of the paper is written on the basis that ($\sin M = \pm 1$).

It is of interest to compare some of the above results with those for the case of a single conductor far removed from reflecting surfaces. If the wire is excited so as to bear standing waves of I r.m.s. amperes maximum value the radiated power is:

$$P_3 = 30I^2\Bigg[0.5772 + \log_\epsilon(2L) - Ci(2L) - \cos^2 M - \cos^2(L - \text{M}) + 2\cos M\cos(L - \text{M})\frac{\sin(L)}{(L)}\Bigg] \text{ watts.} \tag{3}$$

If the wire is "terminated" so that there are no reflections from the ends a uniform current of I r.m.s. amperes may be assumed to exist along the wire. In this case the power radiated is:

$$P_4 = 60I^2\left[0.5772 - 1 + \log_\epsilon(2L) - Ci(2L) + \frac{\sin 2L}{2L}\right] \text{ watts.} \tag{4}$$

An Efficient Miniature Condenser Microphone System *

By H. C. HARRISON and P. B. FLANDERS

It has been shown recently that microphones and contiguous amplifiers distort the sound field in which they are placed by reason of their size and the cavity external to the diaphragm of the microphone. For frequencies such that the size is large compared to the wave-length of perpendicularly incident sound, reflection causes the actuating pressure to be double that which would exist in the undisturbed field. If the direction of the incident sound be along the plane of the diaphragm, the increase of pressure due to reflection is not as great; but there may be a substantial reduction in effective pressure due to differences in phase across the diaphragm. In addition, cavity resonance produces an increase of pressure at frequencies usually within the working range of the microphone.

This paper describes a laboratory model of a Wente-type condenser microphone of high efficiency and an associated coupling amplifier which are of such small size that reflection and phase-difference effects are of negligible importance within the audible frequency range; while the cavity is so proportioned that its resonance effect is an aid rather than a detriment to uniformity of response in a constant sound field.

SEVERAL writers [1] have recently called attention to the fact that a microphone distorts the sound field in which it is placed by reason of its size and the cavity external to the diaphragm. The distortion due to size was first mentioned by I. B. Crandall and D. MacKenzie in 1922.[2] It is a function of the direction of the sound with respect to the diaphragm.[3] The distortion due to cavity resonance is substantially independent of direction and depends mainly on the relation between the dimensions of the cavity and the wave-length of sound.

If a microphone were to be designed so that it would respond uniformly to sound coming from any direction, it is apparent that first the size would have to be diminished to such an extent that reflection and phase-difference effects became negligible. Secondly, the cavity would either have to be eliminated entirely [4] or else be so proportioned that resonance occurred at frequencies above the resonance frequency of the diaphragm, where the response of the latter was diminishing. Such mutual compensation is possible in a small microphone and the effect is substantially independent of the direction of sound.

---

* Presented before Acous. Soc. Amer., New York, N. Y., May 3, 1932.

[1] A. J. Aldridge in *P. O. E. E. Jour.*, Oct., 1928, pp. 223–225; S. Ballantine in *Phys. Rev.*, Dec., 1928, pp. 988–992; W. West in *I. E. E. Jour.*, 1929, pp. 1137–1142.

[2] *Phys. Rev.*, March, 1922.

[3] L. J. Sivian in *B. S. T. J.*, Jan., 1931 pp. 96–116.

[4] S. Ballantine in "Contributions from the Radio Frequency Laboratories," No. 18, April 15, 1930.

This paper will discuss the factors relating the dimensions of a condenser-type microphone system with the several types of field distortion, and will describe a miniature system designed to practically eliminate such distortions over a wide frequency range.

Diffraction of Sound Around an Obstacle

The shapes of condenser microphones alone or in association with amplifiers are so irregular that it is impossible to calculate their effect as diffractors of sound waves. It will suffice, however, to assume some regular shape approximating actuality, for which calculations can be made. The diffraction effects so obtained for the regular shape will be substantially the same as those caused by the actual irregular shape, provided their areas projected on the plane of the incident sound-waves be equal. Cavity resonances, of course, may be quite different in the two cases. But these can be treated as separate effects, and will be so considered in a later section of the paper. In this section, only those disturbances of the sound field, caused by the smooth envelope of the microphone alone or with its amplifier, will be considered.

The case of diffraction of plane sound waves around a spherical

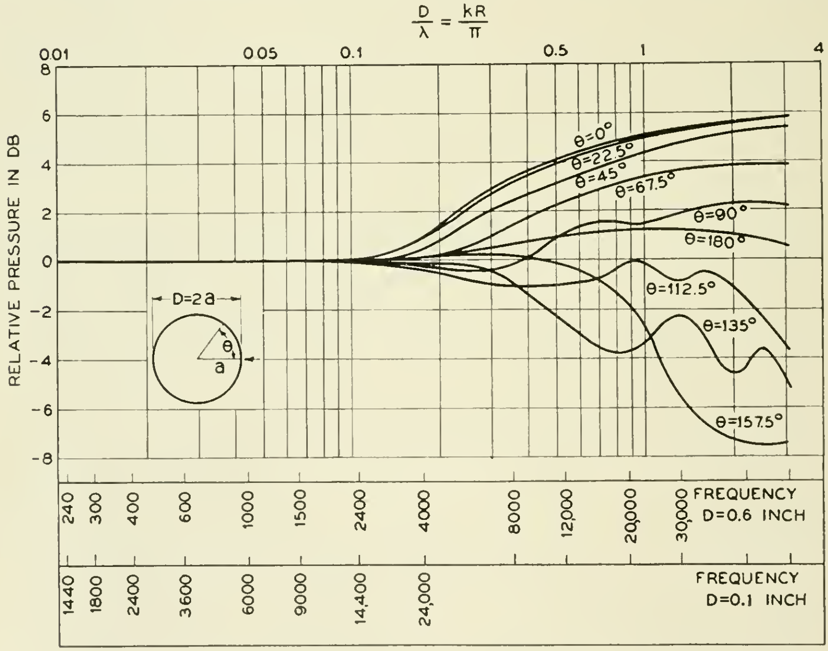


Fig. 1—Variation in pressure on rigid sphere due to diffraction of plane sound waves.

obstacle is amenable to mathematical treatment, and was considered theoretically by Raleigh. Quantitative consideration of the effect at a point on a sphere directly in line with the oncoming sound has been given by S. Ballantine.[1] Fig. 1 shows the effect for other polar angles around the sphere, as computed from Ballantine's equation. Of particular theoretical interest is the curve for a polar angle of 180°; that is, the point that should be most completely "shadowed" from the sound. Actually, no shadowing effect appears, the pressure remaining substantially equal to that of the undisturbed field. This case is analogous to that where diffraction of light causes a bright spot to appear in the center of the shadow of a circular disk. The area of this acoustic bright spot is small, as may be seen by the pronounced shadowing of a point only $22\frac{1}{2}°$ away. Because of its small area, it is impractical to make use of the effect in microphone design.

EFFECT OF PHASE-SHIFT IN A PLANE SOUND WAVE TRAVELING ALONG THE PLANE OF THE DIAPHRAGM

In Appendix I is given an approximate calculation of the reduction in effective pressure on a circular diaphragm, due to the change in phase of a progressive plane wave traveling across it. The effect is plotted in Fig. 2. As might be expected, the reduction becomes

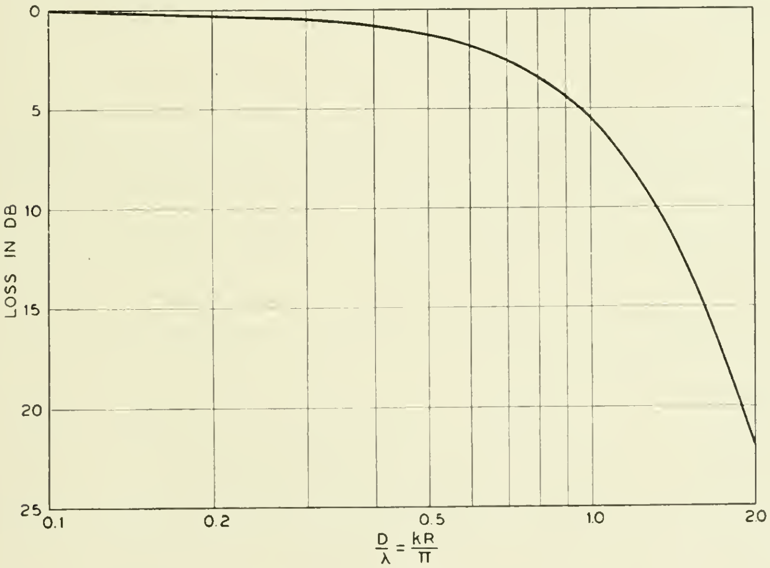


Fig. 2—Loss in effective pressure due to phase-shift in plane sound waves traveling across circular diaphragm.

---

[1] Loc. cit.

considerable when the diameter of the diaphragm exceeds the wave-length. Since in normal use, a large proportion of the sound actuating the microphone comes in a transverse direction, the distortion due to this phase-cancellation effect may often be more serious than the distortion due to diffraction.

Cavity Resonance

The effect of a cylindrical cavity on the pressure actuating a diaphragm is given by equation (5) in Appendix II and is shown graphically in Fig. 3. It is apparent that the larger the ratio of diameter to wave-length, the smaller is the resonance effect; and that the greater the ratio of a given diameter to the depth of the cavity, the higher the frequency at which such resonance occurs. There is, of course, more than one resonance. However, for the higher resonances there is a greater ratio of diameter to wave-length with a consequent increase in damping; so that the resonance effect is less pronounced. Compared to the primary resonance, the secondary resonances are, as a rule, of negligible importance.

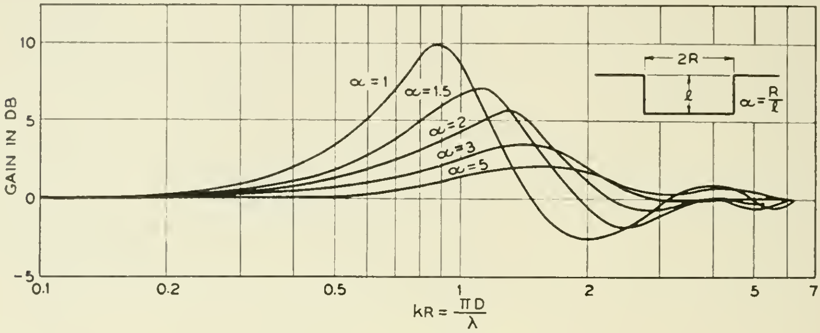


Fig. 3—Gain in effective pressure on circular diaphragm due to cavity resonance.

Acoustic and Electrical Considerations Affecting the Size of a Condenser Microphone

If diffraction effects for all angles of incidence are to be negligible in a frequency range extending, for example, to 15,000 c.p.s., it is apparent from Fig. 1 that the diameter of the microphone should not be greater than about a tenth of an inch. Assuming that all other dimensions and characteristics remain fixed, including the resonance frequency of the diaphragm and the ratio of dead to active capacity, it can be shown that as the diameter of a condenser microphone diaphragm is decreased, the ratio of generated voltage to actuating pressure remains constant. The diameter can not be decreased indefinitely, however, because of the limitations of amplifiers with which the condenser microphone must of necessity be closely associated. The resistor for feeding the polarizing voltage and the grid biasing resistor must be increased as the diameter of the diaphragm (and consequently the active capacity) is reduced, in order that, at low frequencies, the same proportion of generated voltage may get to the grid of the amplifier tube. They must not be made too large, however, for then the voltages due to thermal agitation [5] in the resistors will become comparable to the signal voltage. Similarly, the capacity of the input leads, vacuum tube and resistors will become comparable to the lowered capacity of the microphone, with a consequent reduction in signal voltage at the grid, over the whole frequency range. The actual limit in reducing microphone size can not be defined accurately; but it is definitely greater than one-tenth of an inch.

Further consideration of the diffraction problem, however, shows that such an extremely small size is not really necessary in order to practically eliminate sound-field distortion. When sound is picked up indoors at some distance from the source, the directly incident sound contributes much less to the microphone output than does the reflected sound arriving from other angles of incidence.[3] The greater part of the effective actuating pressure comes then at polar angles in the vicinity of 90°, where the diffraction effect is much less pronounced. If 90° be taken as an effective average angle, it is seen from Fig. 1 that the diameter of a microphone need only be reduced to about six-tenths of an inch in order to make this type of distortion negligibly small.

If sound is picked up out-of-doors, or indoors near the source, the directly incident waves predominate over the reflected waves. For this case it will suffice to place the microphone so that the sound arrives at a polar angle of 90°; and a six-tenths inch diameter is still sufficiently small.

Inspection of Fig. 2 shows that the phase-difference loss is about 1 db at 10,000 c.p.s. for this diameter; so that six-tenths of an inch can safely be chosen as an acceptable design value.

The Amplifier

Consideration has now been given to the problem of reducing the size of a microphone to such an extent that it would not appreciably disturb the sound-field. But all such labor is in vain if the size of the coupling amplifier be not correspondingly decreased, since, for efficient operation, the high impedance of the microphone necessitates a close spacial coupling between the two.

---

[5] J. B. Johnson, *Phys. Rev.*, July, 1928, pp. 97–109.

[3] Loc. cit.

Use of a special miniature type vacuum tube affords the possibility of materially reducing the size of the coupling amplifier. It is but slightly greater in diameter than the microphone just described, the inter-electrode capacities are quite low, and the plate resistance is not too high. This tube, with the necessary coupling resistors and a small stoppage condenser are placed within a cylindrical metal tube of about 0.8 inch diameter, to one end of which is attached the condenser microphone. From the other end of the cylinder extends a shielded cable along the axis of which runs the plate lead from the vacuum tube. This cable is constructed so that the capacity of the lead to ground is small. Surrounding the lead are two filament supply conductors and a conductor for supplying polarizing voltage to the microphone. The impedance between these and ground being very small, their capacity to ground may be as large as desired. Of course, the longer the connecting cable, the lower must be the capacity of the plate lead per unit length. The second stage, to which the cable runs, may be large, and as distant from the miniature first stage as is consistent with these lead capacity requirements.

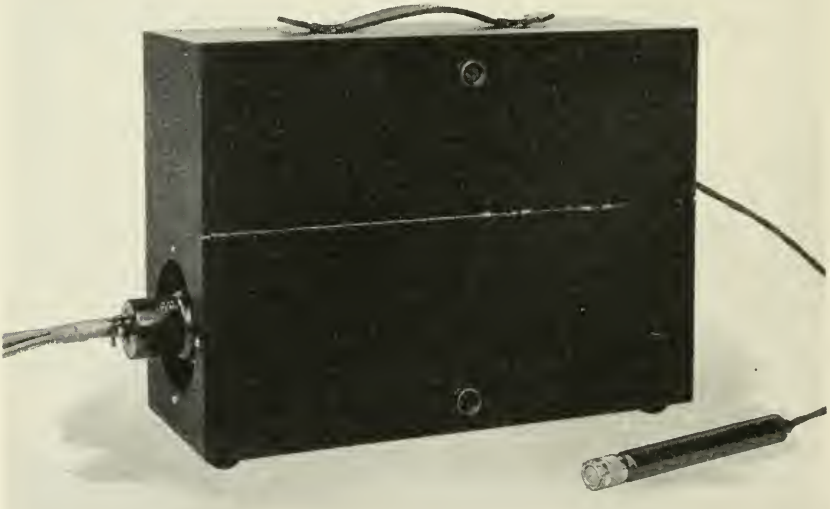


Fig. 4—View of miniature microphone and attached miniature coupling amplifier, together with second amplifier stage.

The miniature microphone and the miniature vacuum tube with necessary coupling resistors and a stoppage condenser are contained in the metal tube shown in the right foreground of Fig. 4. The dimensions of the container tube are about 7.5 inches long and 0.8 inch in diameter. The carrying case houses the second stage amplifier together with other accessories.

CONCLUSION

For the same frequency range, the efficiency of this miniature microphone as determined by a thermophone calibration is about 2.5 db greater than that of the 394W type, because of a lower proportion of dead to active capacity. In combination with a well designed amplifier, the efficiencies as determined by the voltages on the grid of the tube are about equal, most of the proportional contribution to dead capacity in the one case coming from the amplifier, and in the other case from the microphone.

In this miniature condenser microphone, the diaphragm is tuned higher than is the 394W, so that the efficiency in association with the amplifier is about 3 db lower than that of the 394W. The resonance of the external cavity gives a maximum lift in response of about 3 db. The microphone as a whole responds uniformly up to 10,000 c.p.s. The diffraction and phase-difference effects are negligible up to that frequency.

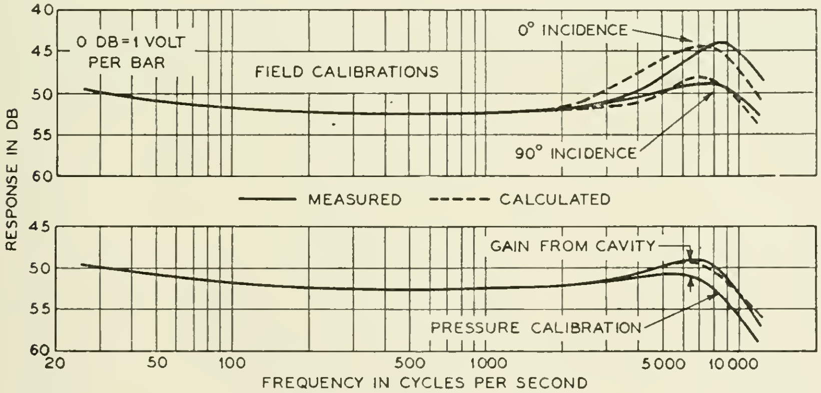


Fig. 5—Pressure and field calibrations of miniature condenser microphone, showing cavity effect.

Curves showing the constant pressure and constant field[3] calibrations of this microphone are given in Fig. 5.

The work on this miniature microphone system has so far been essentially of a research nature, and its use has been directed primarily to fundamental laboratory problems in the field of sound measurements. The necessary work toward commercialization has not been effected, in view of the high quality performance and certain practical advantages of available types of microphones.

---

[3] Loc. cit.

APPENDIX I

For a stretched membrane, the equation expressing the statical deflection ξ as a function of the radius r, for a symmetrical distribution of pressure p, is [6]

$$p = -\frac{\tau}{r}\frac{d}{dr}\left(r\frac{d\xi}{dr}\right), \tag{1}$$

where τ is the tension coefficient. Two integrations give

$$\xi = \frac{1}{\tau}\int_r^R \left[\frac{1}{r}\int_0^r prdr\right] dr, \tag{2}$$

where R is the bounding radius. The central displacement is evidently

$$\xi_0 = \frac{1}{\tau}\int_0^R \left[\frac{1}{r}\int_0^r prdr\right| dr. \tag{3}$$

Now, considering the case of a progressive plane wave

$$p = P\cos(\omega t - kx)$$

$\left(\text{where } k = \frac{2\pi}{\text{wave-length}}\right)$ traveling across this membrane, an expression can be computed for the average pressure on the boundary of the circle of radius r. From Fig. 6 it is clear that

$$x = R - r\cos\theta$$

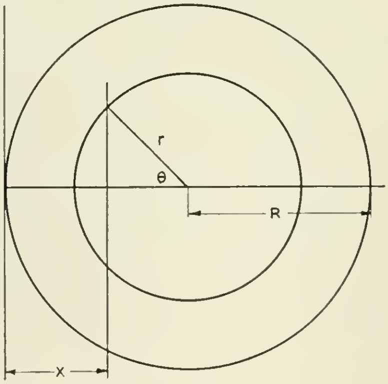


Fig. 6—

---

[6] See Lamb's "Dynamical Theory of Sound," § 54.

and hence that

$$\begin{aligned} p &= P \cos\left[(\omega t - kR) + kr\cos\theta\right] \\ &= P\left[\cos(\omega t - kR)\cos(kr\cos\theta) - \sin(\omega t - kR)\sin(kr\cos\theta)\right]. \end{aligned}$$

The average pressure is

$$\begin{aligned} p_{\text{av}} &= \frac{1}{\pi}\int_0^\pi p\,d\theta \\ &= \frac{P\cos(\omega t - kR)}{\pi}\int_0^\pi \cos(kr\cos\theta)d\theta \\ &\qquad - \frac{P\sin(\omega t - kR)}{\pi}\int_0^\pi \sin(kr\cos\theta)d\theta. \end{aligned}$$

Here the second integral is zero, while the first may be written

$$\begin{aligned} p_{\text{av}} &= \frac{2P}{\pi}\cos(\omega t - kR)\int_0^{\pi/2}\cos(kr\cos\theta)d\theta \\ &= \frac{2P}{\pi}\cos(\omega t - kR)\int_0^{\pi/2}\cos(kr\sin\theta)d\theta \end{aligned}$$

and is readily identified as a Bessel function.

$$p_{\text{av}} = P\cos(\omega t - kR)J_0(kr). \tag{4}$$

Substitution of (4) in (3) gives an approximation to the total effect (on the central displacement) of these average pressures acting over the whole diaphragm.

$$\xi_0 = \frac{P\cos(\omega t - kR)}{\tau}\int_0^R\left[\frac{1}{r}\int_0^r J_0(kr)rdr\right]dr. \tag{5}$$

Now

$$\int_0^r J_0(kr)rdr = \frac{r}{k}J_1(kr)$$

and

$$\int_0^R \frac{1}{k}J_1(kr)dr = \frac{1}{k^2}\left[1 - J_0(kR)\right]$$

so that

$$\xi_0 = \frac{P\cos(\omega t - kR)}{k^2\tau}\left[1 - J_0(kR)\right]. \tag{6}$$

The displacement when R is small compared to the wave-length, i.e. $k = 0$ is

$$\xi_0' = \frac{PR^2\cos\omega t}{4\tau}. \tag{7}$$

The ratio of (6) to (7) in amplitude is

$$\left|\frac{\xi_0}{\xi_0'}\right| = \frac{4[1 - J_0(kR)]}{(kR)^2}, \tag{8}$$

which is plotted as db loss *vs.* kR in Fig. 2.

There are two approximations involved in this analysis, both involving equation (3). One is that for p (which should be constant for a given r), can be taken the average of the actual pressures around the circle of radius r. The other is that the shape of the static deflection curve represents the actual shape up to the highest frequencies of interest in (8).

APPENDIX II

Assuming (1) that the air particles, in the plane of the entrance to the cavity, all move in phase with equal velocities v_1 which are normal to that plane, and (2) that the impedance per unit area of the microphone diaphragm is large compared to ρc, where ρ is the density of air and c is the velocity of sound, the following three relations hold: From the theory of plane wave propagation in a tube,

$$p_2 = p_1 \cos kl - i\rho c v_1 \sin kl \tag{1}$$

where p_1 is the pressure in the plane of the entrance to the cavity of depth l and p_2 is the pressure at the diaphragm. Also, the input impedance per unit area of this closed cylindrical tube is

$$\frac{p_1}{v_1} = -\,i\rho c \cot kl. \tag{2}$$

Now p_1 is equal to the pressure P that would exist at the opening if the air particles were held stationary, diminished by the drop in pressure due to their motion and the consequent radiation from the opening. In symbols,

$$p_1 = P - \rho c(a + ib)v_1, \tag{3}$$

where $(a + ib)$ is the radiation impedance coefficient given by Raleigh [7] for the case of a circular piston in an infinite wall:

$$\begin{aligned} a &= 1 - \frac{J_1(2kR)}{kR}. \\ b &= \frac{8kR}{\pi}\sum_{n=1}^{n=\infty}(2n+1)\left(\frac{2^n\,\underline{|n}}{\underline{|2n+1}}\right)^2(-\,4k^2R^2)^{n-1}. \end{aligned} \tag{4}$$

---

[7] "Theory of Sound," Vol. II, § 302.

Elimination of p_1 and v_1 from the first three equations gives

$$\left|\frac{P}{p_2}\right| = \sqrt{(\cos kl - b \sin kl)^2 + a^2 \sin^2 kl} \tag{5}$$

as the expression for the change in acoustic pressure on a diaphragm, caused by the entrance cavity.[8] It is plotted as db gain in Fig. 3.

Similar calculations have been given by other writers.[3, 9]

---

[8] Numerical values of the functions a (kR) and $b(kR)$ are given in Crandall's "Theory of Vibrating Systems and Sound," Fig. 19, p. 172.

[3] Loc. cit.

[9] W. West, in *Jour. I. E. E.*, April, 1930.

Wire Communication Aids to Air Transportation *

By H. H. NANCE

RAPID development of air transportation in this country has continued through the past few years and today established routes connect nearly all important cities. The route mileage of the airways in the United States as shown in Fig. 1 totals over 30,000 miles. Regularly scheduled transport service is given on practically all of these routes and considerable use of them is also made by military and private planes. Statistics relating to service of air transport companies seem particularly significant. The United States Department of Commerce reported approximately 42,800,000 miles flown in passenger, mail and express service on domestic scheduled lines in 1931, an increase of 35 per cent over the preceding year and more than a fourfold increase since 1928. In the same three-year period passengers carried increased ninefold, reaching a total of around 470,000 in 1931. Along with this growth safety has been increased as indicated by the respective 1928 and 1931 reports of 250,000 and 750,000 miles flown per accident. Reasonable regularity of schedules on air transport lines also has been maintained, the ratio of miles actually flown to scheduled miles last year being in the order of 92 per cent.

Communication facilities have been an important contributing factor to all this development and improvement. It was recognized early that fast and reliable communication would be needed in connection with any extensive development of air transportation. Communication with planes in flight was an obvious requirement and this could be provided only by radio. For land service, however, experience has indicated that wire facilities best meet the general requirements. This paper describes the wire communication facilities in general use today, both by the Government and by transport companies, as aids to air transportation.

Principal airways have been established largely through Federal aid. In addition to marking and lighting airways the Airways Division of the Department of Commerce had provided up to April 1, 1932, 67 radio telephone stations at approximately 200-mile intervals as indicated in Fig. 1, to be used for broadcasting weather reports and similar

---

* Presented at Summer Convention of A. I. E. E., Cleveland, Ohio, June 20–24, 1932. Published in abridged form in *Elec. Engg.*, July, 1932.

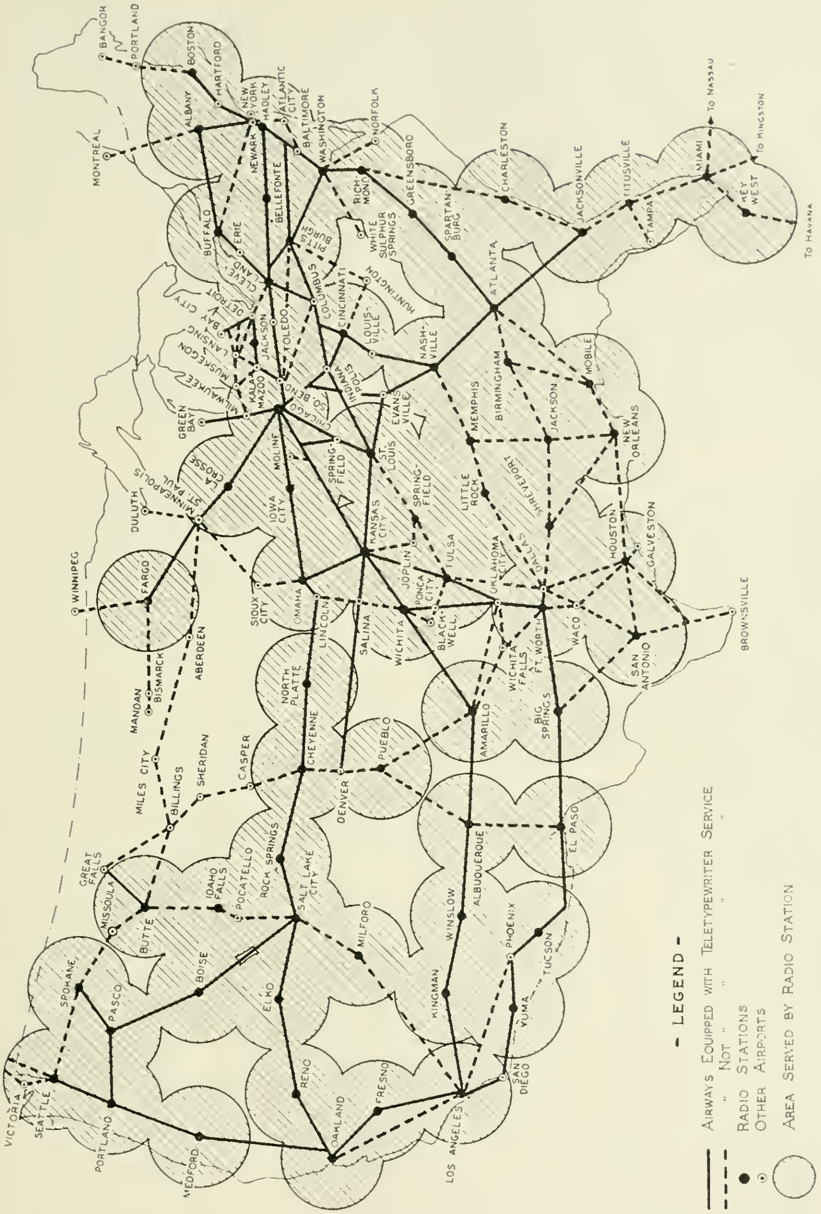


Fig. 1—Airway map of the United States showing routes equipped with teletypewriter service.
]

information to planes in flight and for transmitting directive radio beacon signals for enabling planes to keep on the course. In conjunction with these services it had contracted for 24-hour teletypewriter service along 13,000 miles of main airways connecting some 250 stations, principally for the purpose of transmitting weather reports and to assist in dispatching the planes. All of these facilities have been made available without cost to aircraft operating companies and others using the airways.

In addition to this communication service contracted for by the Government, approximately 5,000 miles of teletypewriter circuit are used daily in furnishing private wire communication service to a number of transport companies for transmitting information pertaining to the operation of their own lines. Routes on which these facilities are furnished to the Government and transport companies are indicated by heavy lines in Fig. 1.

When the first air mail service was established, radio telegraph was introduced as a means of point-to-point communication along the New York-Chicago-San Francisco airway route. At each radio station meteorological data were collected from surrounding points by means of long distance telephone and telegraph and these data were exchanged periodically through the day with the other stations over the radio telegraph.

With a rapid expansion in air transport service foreseen it was apparent there would be a large increase in communication requirements not only to equip new routes but to handle increased volume on existing routes. There was the definite requirement for radio telephone communication with planes which would need a number of the radio channels allotted to this service. Considering these factors and the geographic and other conditions applying to probable development of air transportation in the United States, it seemed that regular point-to-point service served by radio telegraph could be provided more satisfactorily in another way.

Arrangements were made in 1928 for teletypewriter communication services at several points connecting radio stations with their local weather bureau offices in order to expedite the delivery of weather reports and other traffic handled by radio telegraph. Shortly afterward, a teletypewriter system was installed on the New York-Cleveland route connecting the Department of Commerce and Weather Bureau stations at Hadley Field, Stelton, N. J., and Cleveland, Ohio, and a number of intermediate points. This type of service seemed ideally fitted for use in weather reporting and plane dispatching and has been extended not only to replace the service furnished by the radio telegraph system but also to provide for communication requirements on other routes.

Teletypewriter service offers the advantages of simultaneous communication with any desired number of stations, the communications being automatically recorded on machines at each point. A message using code or abbreviations, if desired, can be sent instantly without the necessity of calling in or checking with the receiving stations; thus the immediate attention of only the sending operator is required. Automatic recording reduces the possibility of human error and permits the most efficient use of operating personnel with resulting savings in labor. Furthermore, as contrasted with radio, this system, utilizing wire transmission, is not so subject to variations in meteorological conditions; it is thus more dependable, and has the advantage that it can be readily extended to handle large volumes of business. This system also is well adapted for carrying on administrative and other work as well as for weather reporting and plane dispatching.

Transmission of Weather Reports

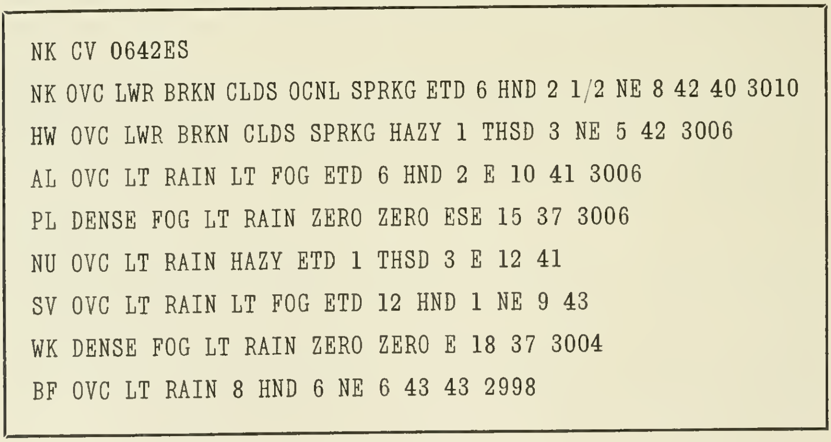
Material progress has been made in reducing the effect of weather hazards to air transportation, through the service rendered by the Department of Commerce and United States Weather Bureau in the collection and dissemination of weather reports supplemented by other reports collected by individual transport companies from planes in flight. For this service a system of practically continuous reporting and forecasting for areas along air routes has been developed and weather observations have been extended to include data of particular benefit to air navigation.

The teletypewriter networks furnished the Department of Commerce are devoted largely to this purpose and in conjunction with its radio telephone broadcasting service are the means for providing to pilots information relating to existing conditions and forecasts for both general and local areas.

Twelve selected Weather Bureau airport stations located at strategic points in the country's airway network prepare summaries of weather conditions in their own areas and make area forecasts every three hours based on data collected over the Department of Commerce circuits from connected airway stations and over commercial telegraph lines from other reporting points. These summaries and forecasts are then transmitted over the teletypewriter circuits and made available to all airway stations.

While the forecasts include predictions as to storm developments or movements, conditions in specific localities are often likely to change rapidly and it has been necessary to provide additional reports along the air routes on an hourly basis in order to keep pilots continuously advised of conditions likely to be encountered. Consequently, the airway keepers and Weather Bureau observers at the various teletypewriter stations make local observations of general weather conditions, ceiling height, visibility, wind direction and velocity, temperature, and barometric pressure, every hour. These data are then sent by teletypewriter and automatically recorded at all points in accordance with a predetermined schedule, which is coordinated with the broadcast schedule of radio stations. Since pilots will tune in on particular stations at definite times to obtain reports in accordance with the broadcast schedule, it is important that the schedules be closely adhered to. The following is an illustration of scheduled weather reporting along the Newark-Cleveland route.

At 42 minutes past the hour the observers will begin typing their observations on the circuit beginning with the Newark station, followed successively by Hadley, Allentown, Park Place, Numidia, Sunbury, Winkleblech, Bellefonte, Kylertown, Greenwood Club, Brookville, Mercer, Parkman and Cleveland, with practically no interval between the completion of the report from one station and the beginning of a report from the next. When the Cleveland weather observer has completed typing his report a complete record of weather conditions at all points on the circuit will appear on the teletypewriter tape at each individual station and in the radio broadcasting stations located at Hadley, Bellefonte and Cleveland. Fig. 2 shows a portion of an actual tape record of an hourly report along the Newark-Cleveland route which includes the stations between Newark and Bellefonte, Pa. For convenience the tape has been cut to show one station report on



```
NK CV 0642ES
NK OVC LWR BRKN CLDS OCNL SPRKG ETD 6 HND 2 1/2 NE 8 42 40 3010
HW OVC LWR BRKN CLDS SPRKG HAZY 1 THSD 3 NE 5 42 3006
AL OVC LT RAIN LT FOG ETD 6 HND 2 E 10 41 3006
PL DENSE FOG LT RAIN ZERO ZERO ESE 15 37 3006
NU OVC LT RAIN HAZY ETD 1 THSD 3 E 12 41
SV OVC LT RAIN LT FOG ETD 12 HND 1 NE 9 43
WK DENSE FOG LT RAIN ZERO ZERO E 18 37 3004
BF OVC LT RAIN 8 HND 6 NE 6 43 43 2998
```

Fig. 2—Teletypewriter tape with portion of weather sequence report.

each line. First the starting time, 0642 E. S., which is 6:42 a.m. Eastern Standard Time, is shown. Each reporting station in sequence then gives its code letter or letters and follows with a report of its observations. An interpretation of the report from the first station is "Newark, overcast, lower broken clouds, occasional sprinkling, estimated ceiling height 600 feet, visibility 2½ miles, wind velocity 8 miles per hour, direction northeast, temperature 42°, dew point 40°, barometric pressure 30.10 in." The time of actual transmission for all 14 stations, Newark to Cleveland, is generally about four minutes.

At 50 minutes past the hour the three radio stations will interrupt the beacon signals and broadcast the reports just received. Hadley station transmits the weather sequence received from stations between Newark and Bellefonte; simultaneously, the Bellefonte radio station transmits the entire sequence received from all points between Newark and Cleveland, and the Cleveland radio station broadcasts reports from points between Bellefonte and Cleveland. All three radio stations include in the sequence, reports of weather at Cleveland and New York. The range beacons are not interrupted for these reports for longer than two minutes, and if the reports require a longer period the beacon signals are restored for one minute and again interrupted to complete the reports.

Based upon the information obtained through the sequence collections, the airway weather reporting station retransmits, generally by teletypewriter, hourly weather reports to the various airway operating companies' offices in that vicinity. Airway companies maintain various arrangements for posting the weather information for the convenience to pilots. Some companies post the information on a series of boards of different color arranged in geographic sequence to represent different airway routes, each board indicating a particular point on that route.

An experimental service involving the transmission of weather summaries in map form has been tried out recently at Kansas City, Chicago, Cleveland, Newark and Washington. A separate circuit equipped with page teletypewriters at each of these points was provided for this purpose. The weather maps were prepared at Kansas City and Cleveland every three hours. A typical map, the notations of which were transmitted over the circuit and directly printed by the teletypewriter, is shown in Fig. 3, and the following describes briefly the methods used.

Two special airways maps, ordinary letter width, have been printed, one map covering the section of the country east of the Mississippi River and another the section west to the Rocky Mountains. The maps are printed in ink which permits hectographic reproduction. The

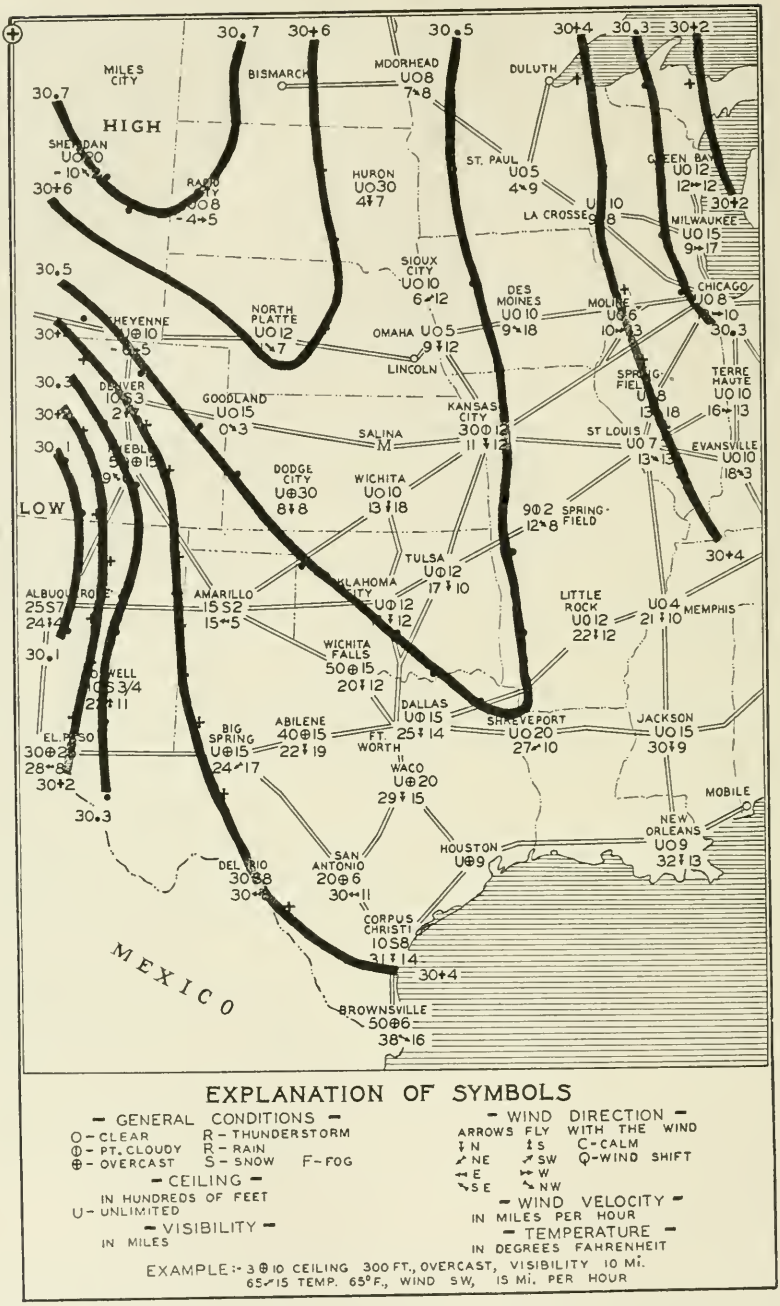


Fig. 3—Weather map, the notations of which were transmitted and directly printed by the teletypewriter system.

airways, principal airports, and cities are shown on the map and in the upper left corner is a small circle used as a coordinating point.

At a scheduled time the operator at Kansas City or Cleveland inserts in a teletypewriter equipped for perforating tape a copy of the map on which the latest weather information has already been typed including general state of weather, ceiling height, visibility, temperature, wind direction and velocity and barometric pressure for each point and isobars connecting points of equal pressure. The sending operator then types the identical symbols, letters, and figures directly over the corresponding ones on the map inserted in his machine, thus making a complete record on perforated tape. On schedule a blank map is inserted in the teletypewriter at each receiving point and positioned so that a type bar will strike the map within the small coordinating circle. The sending operator then releases the tape and the signals transmitted over the circuit reproduce on the map at the receiving stations data similar to those on the original map at the sending station.

The map data are sent in sequence from the two transmitting stations and after they have been received on the map forms a number of duplicate copies can be run off immediately and the two maps fitted together if desired. The maps are then available to pilots at each of the respective airports.

Complete reports of weather are generally maintained by transport companies in dispatching offices. On some lines two-way short-wave radio telephone equipment has been provided for communicating with planes and periodic contact is maintained during flight. In this way pilots report their positions directly to dispatchers and in addition supplementary weather data are usually exchanged, particularly in respect to local ceiling heights and conditions in the upper air strata.

Plane Dispatching and Other Service

Teletypewriter circuits furnished to air transport companies are used principally for dispatching planes and handling the many traffic matters usual to this type of service. Plane movements including reports of position in flight are transmitted over the teletypewriter system and recorded at various offices. The reports of positions, in many cases, are given by pilots over short-wave radio telephone where this type of equipment has been provided.

To facilitate position reporting some of the companies have superimposed a system of rectangular coordinates over a map of the course cutting the territory into squares or rectangles 10 to 20 miles on a side. The coordinates are numbered so that the pilots and dispatchers can readily establish the location of the plane. The dispatchers generally maintain a typewritten, chronological log of position reports from each plane in the air. Bulletin boards are also used, marked with the stations along the route and with spaces for filling in data such as plane number and license, name of pilot, time of arrival and departure at each station and final destination.

A considerable volume of information is required to be transmitted in connection with the handling of traffic on large lines. This usually consists of data as to reservations, number of passengers and amount of mail and express carried, connections to be made, and arrangements at terminals. Supplementary instructions to pilots and many administrative matters requiring prompt handling are also transmitted.

Although the airways teletypewriter circuits furnished the Department of Commerce are used mostly for handling weather reports, considerable information is also transmitted over them relating to departure and arrival of planes and their position in flight. Upon request the Department of Commerce will send over its teletypewriter system the license number of a plane, the station from which the plane is departing, its time of departure, and its destination, to stations along the route of the flight. Stations on the route knowing approximately the time the plane will be due watch for it and record the actual time the plane passes so that other stations may be informed.

Teletypewriter Circuit Layout

Teletypewriter networks furnished by the Bell System for service along airways are composed of some 30 separate circuits. Circuit mileage of the longest is about 2,000 miles and of the shortest, 200 miles. The longer circuits generally connect 15 to 20 intermediate stations. Since airways naturally follow direct air lines the intermediate airway stations are often located at points considerable distances from main communication lines, which, generally, are constructed along routes connecting the industrial and more populous centers, due regard being given to topographic and other conditions. At the larger airports such as Newark and Cleveland, local teletypewriter circuits are also provided to connect the Department of Commerce station with the offices of the various transport companies, the post office, and weather bureau. Automatic transmission equipment is provided so that information received on one circuit can be retransmitted over one or more other circuits if desired.

A layout diagram of a typical circuit is shown in Fig. 4. Facilities in the New York-Cleveland long distance cable are used for establishing the main links totaling 515 miles. Repeater stations on the cable route located approximately every 50 miles afford convenient points from which branch circuits are extended to the intermediate airway stations. The several branch circuits are of the grounded open-wire type and total 331 miles. A total circuit mileage, therefore, of 846 miles is required in this case for connecting all stations along an air route a little over 400 miles long.

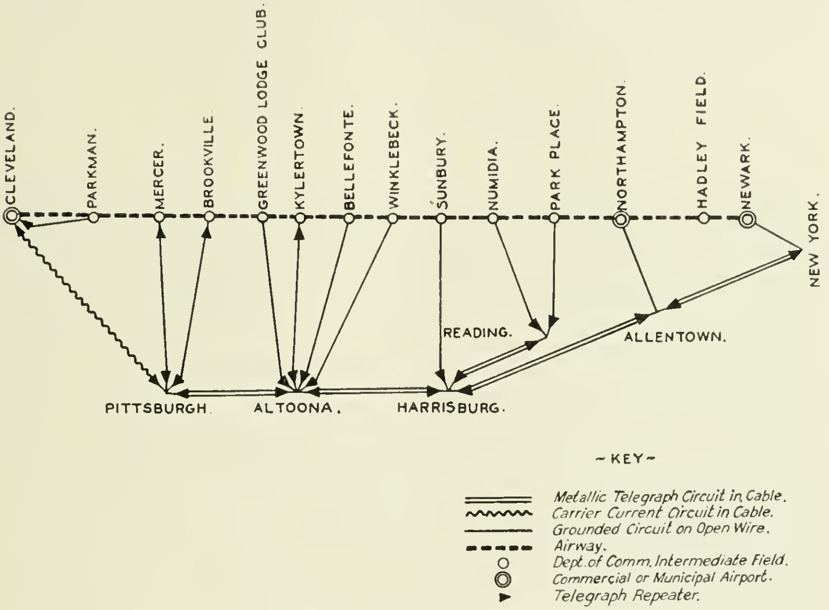


Fig. 4—Layout of typical teletypewriter circuit along airway.

The Newark-Pittsburgh section of the main circuit is operated on metallic telegraph cable facilities, a type particularly adapted to use where stations to be connected to the circuit are spaced at frequent intervals. Between Pittsburgh and Cleveland a channel of a voice frequency carrier telegraph system on cable facilities is used. This type of facility is generally used where stations are located 150 or more miles apart. The longer branch circuits on open wire employ polar transmission with repeaters at both the repeater station and terminal and use two wires, one for each direction of transmission. The shorter branch circuits use one wire with a grounded duplex repeater at the repeater station and a constant d.c. potential at the outlying terminal. Detail descriptions of these various telegraph systems have been given in previous papers.

Cable circuits are less susceptible than open-wire circuits to interference and storm trouble, and where they are available they have been used generally for establishing teletypewriter circuits furnished both the Department of Commerce and the transport companies. At present, over one third of the mileage of these circuits is in cable. Facilities on alternate routes are available to be substituted for the regular circuit in the event circuit trouble develops.

Teletypewriter Equipment

The theory of teletypewriter operation and descriptions of the machines generally used in this country have been given in other papers but are briefly reviewed here in order to describe some of the specific equipment arrangements used in airways service.

The teletypewriter is designed to perform the functions of an ordinary typewriter with added features that permit the typing units of a number of similar machines located at distant points to be controlled by the operation of the keyboard of any one of them. This is accomplished by the translation of the mechanical action of any key in the keyboard unit to electrical impulses arranged in a code and transmitting them over an electrical circuit to the distant machines where the impulses are translated back to the mechanical action of a type bar in the typing unit corresponding to the key struck in the distant keyboard. Electric motors and electro-magnets provide the mechanical power and the means of translation of electrical impulses to mechanical action. It is necessary, of course, that the mechanical action of all of the machines be synchronized. This is provided for by the use of synchronous motors or governed motors regulated to the same speed and a start-stop rather than a continuously rotating system. By the use of the start-stop system the effect of variations in motor speeds is minimized, accurate synchronization being required only during the interval of typing of one letter after which a clutch releases and stops the receiving mechanism momentarily to permit it again to start in synchronism with the sending mechanism. To provide the start-stop feature and sufficient code combinations for the letters and symbols required a seven-impulse code is used consisting of a start pulse, five selecting pulses, and a stop pulse.

Teletypewriters are available to print on an ordinary page or on a narrow strip of tape. Tape machines are generally used in airways service because they are particularly adapted to the handling of short messages and weather sequences where it is generally desirable to rearrange the messages received by cutting and pasting the tape on separate pages to form a continuous weather record for each point. This is preferable to a chronological message record requiring a search through all of the information to obtain the trend of weather at a particular point. The tape machines are also somewhat smaller, less expensive, and more efficient, not requiring the transmission of carriage return and paper feeding signals.

In addition to the ordinary sending and receiving machines supplementary apparatus units may be used so that operators can work at maximum efficiency and the line circuit can be used to its maximum capacity. These units are the perforator, tape transmitter, and reperforator.

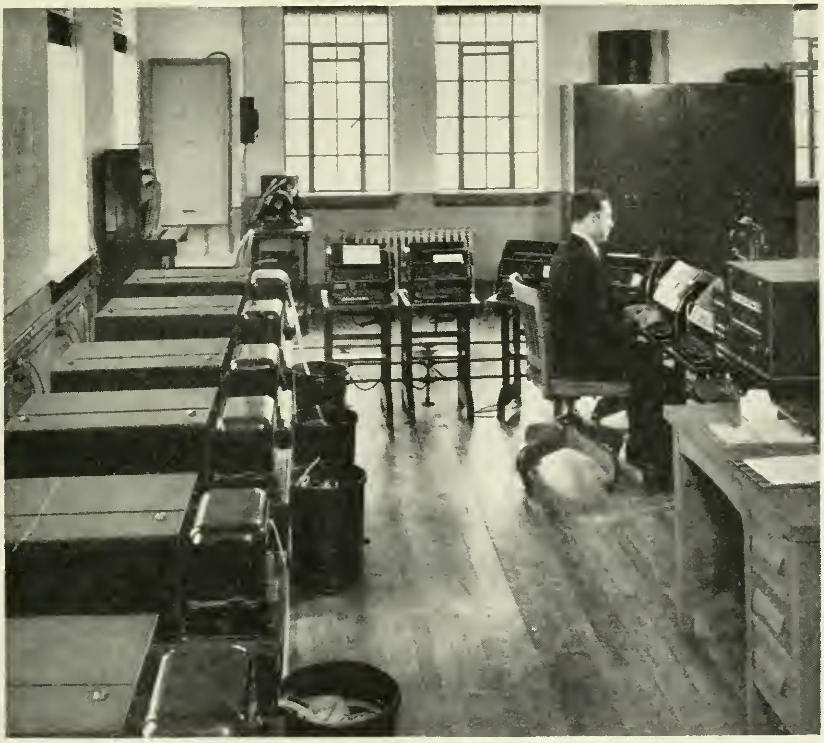


Fig. 5—Department of Commerce airport teletypewriter station.

The perforator is associated with a keyboard and perforates a tape with one to five perforations for each key struck. The tape is run through a tape transmitter which automatically sends electrical impulses to the circuit corresponding to the perforations in the tape and identical to those that would have been sent from the keyboard direct had it been connected to the circuit for normal keyboard sending. The use of the perforator and tape transmitter permits the circuit to be operated at its maximum speed at all times and permits the operator to do other work while the accumulated tape is running through the transmitter. Also the same tape can be run through several tape transmitters and thus be used for sending the message over several circuits.

The reperforator is a receiving device which records the message on a perforated tape similar to that produced by a perforator unit. This permits storing a received message for immediate or subsequent retransmission to other circuits without retyping it.

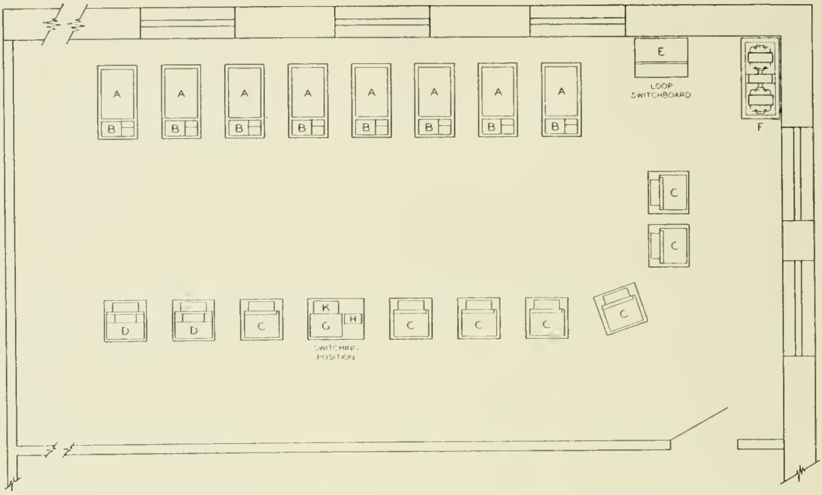


Fig. 6—Teletypewriter equipment layout in large office. *A*, Reperforator; *B*, Transmitter distributor; *C*, Tape teletypewriter; *D*, Page teletypewriter; *E*, Loop switchboard; *F*, Motor generator set; *G*, Radial transmitting board; *H*, Remote control box; *K*, Perforator.

The smaller teletypewriter stations of the Department of Commerce and those of the various transport companies are generally provided with one or two teletypewriters. At the larger teletypewriter stations of the Department of Commerce a special arrangement of the equipment has been provided to permit efficient operation of the teletypewriter circuits from the standpoints of requiring the fewest operators and of obtaining rapid retransmission of messages received on one circuit to one or more of the other circuits as required. A view of an installation is given in Fig. 5 and a typical floor plan arrangement is illustrated in Fig. 6. The apparatus is mounted on tables specially designed for the purpose, and these are usually arranged on the floor in the shape of a U, the units facing inward so that the operators work inside the U.

A separate reperforator (*A*) and tape transmitter (*B*) in addition to a tape teletypewriter (*C*) are provided for each circuit. Messages are received simultaneously on a printed tape by the tape teletypewriter and on a perforated tape by the reperforator. Mounted adjacent to the reperforators are the tape transmitters through which the perforated tape can be run to retransmit immediately a received message to another circuit. The reperforators and tape transmitters can be started and stopped individually from the remote control box (*H*).

A tape perforator (*K*) is provided to perforate tape for messages originating at the local station. The messages can then be sent automatically over the circuit or circuits desired by running the tape through the proper tape transmitters.

All of these units are terminated on cords and plugs at a loop switchboard (*E*) and any unit or combination of units may be connected to any of the teletypewriter circuits which are wired through a number of series jacks in the loop switchboard. A supplementary switching arrangement is provided by radial transmitting board (*G*) equipped with keys and repeating relays. By operating one or more of the keys one of the tape transmitters can be connected quickly to two or more of the teletypewriter circuits through the repeating relays to obtain simultaneous transmission to the circuits connected.

At certain stations page teletypewriters (*D*) are provided for the transmission of weather maps as described previously. This type of machine employs a fixed paper carriage and movable type basket, and accommodates paper up to 8½ inches wide. It has been equipped with a number of special type characters to provide the symbols required on the maps. These symbols, which are shown in Fig. 3, are provided as upper case characters on the teletypewriters in place of fractions and punctuation marks.

The arrangement of equipment described generally permits one operator to attend all of the circuits. The teletypewriters are all located in a fairly small space, which permits one man to observe the incoming messages and operate the control boards, to start and stop the proper transmitter and to relay the messages as required.

RADIO INTERFERENCE

The establishment of teletypewriter stations along the airways brought about the installation of teletypewriter equipment in the same room or in close proximity to short-wave radio receivers, and introduced the problem in specific cases of radio interference caused by the operation of the teletypewriter.

Remedial measures have been designed effectively to reduce this interference, and consist of the use of synchronous motors, rectifiers and specially designed filters, together with the locating of the apparatus and wiring in such a way as to effect a minimum coupling between the teletypewriter and its associated loop and the radio antenna system.

Conclusion

History of air transportation in the past few years indicates that continued growth may be expected, particularly as hazards to flying are mitigated and safety and dependability are recognized by the public. The Government is continuing the extension of airways and weather reporting and other services, and air transport companies are progressing in developing transport business. Fast and reliable communication service has proved the backbone of weather and position reporting and has been a valuable aid in the handling of traffic and other matters relating to air transportation. Teletypewriter circuits used for land service have been found particularly suited to meeting the various requirements involving simultaneous communication with many stations at remote distances. Other wire communication services such as long distance telephone and commercial telegraph have also aided, particularly in reaching points not served by teletypewriter circuits. It is expected wire communication service will continue to be used extensively in connection with air transportation and will be of considerable aid in its future development.

Bibliography

1. *Airway Bulletin* No. 1, September, 1931, issued by the U. S. Department of Commerce.
2. *Air Commerce Bulletin* March 1, 1932, issued by U. S. Department of Commerce.
3. "Metallic Polar-Duplex Telegraph System for Long Small-Gauge Cables," by Bell, Shanck, and Branson, *A. I. E. E. Trans.*, Vol. 44, 1925, p. 316.
   "Voice Frequency Carrier Telegraph System for Cables," by Hamilton, Nyquist, Long, and Phelps, *A. I. E. E. Trans.*, Vol. 44, 1925, p. 327.
4. "Modern Practices in Private Wire Telegraph Service," by R. E. Pierce, *A. I. E. E. Trans.*, Vol. 50, 1931, p. 426.
5. "Police Teletypewriter Communication," by R. E. Pierce, presented at Great Lakes District Meeting, A. I. E. E., Milwaukee, Wis., March, 1932.
6. "Printing Telegraph Systems," by John H. Bell, *A. I. E. E. Trans.*, Vol. 39, Part 2, 1920.
7. "Air Transport Communication," by R. L. Jones and F. M. Ryan, *A. I. E. E. Trans.*, Vol. 49, p. 187.
8. "Aeronautical Communication," by E. Sibley, *Jour. A. I. E E.*, p. 918, November, 1930.
9. "Airplane Flight Aided by Electricity," by C. F. Green, *Electrical Engineering*, August, 1931, p. 654.
10. "Telephone Typewriters and Auxiliary Arrangements," by R. D. Parker, *Bell Telephone Quarterly*, July, 1929.
11. "Teletypewriter Service and Its Present Day Uses," by W. L. Dusenberry, *Bell Telephone Quarterly*, April, 1931.

Abstracts of Technical Articles from Bell System Sources

In January, 1932, a series of seven lectures by representatives of the Bell Telephone System was given before the Lowell Institute of Boston, Massachusetts. The general title of the series was "The Application of Science in Electrical Communication."

The lectures were as follows:

"An Introduction to Research in the Communication Field," by H. D. Arnold, Ph.D., Sc.D., Director of Research, Bell Telephone Laboratories.*

"Research in Speech and Hearing," by Harvey Fletcher, Ph.D., Acoustical Research Director, Bell Telephone Laboratories.

"Transoceanic Radio Telephony," by Ralph Bown, Ph.D., Department of Development and Research, American Telephone and Telegraph Company.*

"Picture Transmission and Television," by Herbert E. Ives, Ph.D., Sc.D., Electro-Optical Research Director, Bell Telephone Laboratories.*

"Talking Motion Pictures and Other By-Products of Communication Research," by John E. Otterson, President, Electrical Research Products, Inc.

"Utilizing the Results of Fundamental Research in the Communication Field," by Frank B. Jewett, Ph.D., D.Sc., Vice President, American Telephone and Telegraph Company, President, Bell Telephone Laboratories.*

"Social Aspects of Communication Development," by Arthur W. Page, A.B., Vice President, American Telephone and Telegraph Company.*

Further Notes on the Detection of Two Modulated Waves Which Differ Slightly in Carrier Frequency.[1] C. B. Aiken. The present paper deals with the analysis of the detection of two modulated waves of slightly different carrier frequency under the conditions that the carrier amplitude of one wave is much smaller than that of the other and that the modulation of the larger wave is low. These conditions apply in determining the interference which arises during the operation of two broadcast stations on the same frequency assignment when the stations are nonisochronous and transmit different programs. A discussion of the characteristics of shared channel interference is given, and it is shown that there are only two important components of this interference, one being the carrier beat note and the other being what has been designated as side band noise. This latter consists of two frequency spectra, one of which is similar to the spectrum of the modulating frequencies of the undesired station but is shifted upward by a constant amount equal to the difference between the carrier frequencies. The other spectrum is of a similar type but is shifted downward in frequency by the same amount.

---

* Published in the April 1932 *Bell Telephone Quarterly.*

[1] *Proc. I. R. E.*, March, 1932.

The Use of Thermionics in the Study of Adsorption of Vapours and Gases.[2] JOSEPH A. BECKER Thermionic emission can be very useful in the study of adsorption phenomena. The primary reason is that very minute amounts of electropositive elements, such as cæsium, barium, or thorium, or electronegative gases, such as oxygen, change the thermionic emission from surfaces of tungsten, platinum, molybdenum, etc., by very large factors and in a characteristic manner. They do this by changing the work function of the surface. This effect, as well as other surface effects, can be best explained by the adion grid theory: The adsorbed particles can exist on the surface either as adions (adsorbed ions) or as adatoms; the adions act like a positively charged, open meshed grid placed very close to the surface. From this theory and the experimental facts it follows: (1) That the ratio of adions to adatoms decreases as the surface concentration increases (Table I); (2) that the work required to remove an adion from the surface increases while the work to remove an adatom decreases as the surface concentration increases; (3) the mean life of an adsorbed particle depends on the surface concentration as well as on the temperature (Table II); (4) the rate of diffusion from the surface into the interior depends upon the temperature and on the amount by which the surface concentration exceeds its equilibrium value. Thermionic experiments show the existence of surface migration and can be used to make a quantitative study of this phenomenon. The techniques involved in these various experiments are described and references given to previous publications.

Electrical Phenomena in Gases.[3] KARL K. DARROW. This treatise of 500 pages is concerned with one of the most important, instructive. and intensively studied fields of modern physics. Its scope embraces, first of all, the elementary processes through the action of which a gas becomes and remains capable of conducting electricity, and acquires and retains other interesting qualities. Such are: "ionization," the detachment of electrons from atoms and molecules by electron-impact, light, impact of positive ions, and other agencies; "excitation," the transfer of atoms and molecules, by these same agencies, into abnormal or excited states in which they possess extra energy and various peculiar powers—the return of excited atoms to their normal condition, with emission of light or with other modes of energy-surrender; "interception," the deflection and slowing-down of electrons and positive ions by collisions with molecules and atoms. After the four chapters devoted to these elementary processes, come four concerned with the drift and diffusion of electrons and more massive ions through dense gases, and with allied topics: material often classified under such names as "mobility," "diffusion," "recombination" and "capture of electrons." The following chapter is assigned to the drifting of ions through dense gases under fields so strong that these ions themselves produce extra ionization, and to the phenomena of "breakdown" which is sometimes sparkover, sometimes the establishment of a discharge such as a glow or an arc. After a chapter on the distortion of electric fields by space-charge, essential to what follows, there comes a treatment of the properties of highly ionized and luminous gases such as the mercury-vapor arc exemplifies, and of space-charge sheaths, as clarified of recent years by newly developed probe-methods. The final chapters are descriptions of the important types of discharge known as self-sustaining glow and self-sustaining arc. The book is thus so arranged as to proceed from the fundamental atomic phenomena, through the intermediate topics of drift and diffusion of ions through gases, to the most intricate discharges. In the later parts, especial attention has been paid to such phenomena as have been made at least partially intelligible in terms of the processes described in the earlier; nevertheless, those which have not been interpreted are given due place and emphasis.

---

[2] *Transactions of the Faraday Society*, March, 1932.

[3] Published by Williams and Wilkins Company, Baltimore, Maryland, 1932.

Electrical Phenomena in Gases is Dr. Darrow's second book. *Introduction to Contemporary Physics* was published by D. Van Nostrand Company in 1926.

Application of Quartz Plates to Radio Transmitters.[4] O. M. HOVGAARD. This paper discusses the disturbing elements encountered in the application of quartz plates to broadcast and aircraft radio transmitters. A general procedure for minimizing such effects is considered from a circuit standpoint as well as in the light of practical experience.

---

[4] *Proc. I. R. E.*, May, 1932.

The degree to which maintenance may affect performance and the necessity for automatic equipment are shown by data obtained in the field. Apparatus and systems which enable the operating staff to meet modern frequency stability requirements by monitoring the emitted carrier are also described.

Tape Armored Telephone Toll Cable.[5] C. W. NYSTROM. Toll cables buried directly in the earth are coming into increased use. Such cables are not installed in the usual clay conduit but are protected by layers of paper, jute, and steel tapes. Complete equipment for laying this cable has been developed.

Some Effects of Topography and Ground on Short-Wave Reception.[6] R. K. POTTER and H. T. FRIIS. This paper contains some results of an experimental study of the effects which ground and ground irregularities have upon short-wave signal reception. The results illustrate the signal strength advantage to be gained in the selection of suitable ground or topographical conditions and show the influence of antenna types, and vertical angle of signal arrival, upon such an advantage. Although the tests were confined to reception, the conclusions are probably applicable in general to the case of transmission. The agreement between measurement data and theory seems to justify the application of plane wave optical theory to the calculation of vertical plane directivity of antennas. Such an application suggests, according to the data obtained, that signals from South America are normally received at much lower vertical angles than those from England.

Western Electric Noiseless Recording.[7] H. C. SILENT and J. G. FRAYNE. The Western Electric method of noiseless recording with the light valve is described. The general principles are discussed, the circuit diagram is explained, and the method of adjusting the device for service described. The photographic characteristics of film are considered, and their application in noiseless recording is shown in some detail.

The Acoustics of Large Auditoriums.[8] S. K. WOLF. Extremely large auditoriums present acoustical difficulties which do not readily yield to the customary methods of analysis and correction. This is illustrated by measurements of the time of reverberation, made in the Madison Square Garden, New York, N. Y., which revealed a considerable discrepancy between theoretical expectations and the times actually measured throughout the frequency range. At 500 cycles, for example, analysis of the auditorium indicated a decay period of 35.5 seconds, whereas the time actually measured by the spark chronograph reverberation meter was only 7.6 seconds. On the basis of the measured time, 47,000 square feet of one-inch rock wool were installed. This material was distributed in a manner calculated to suppress undesirable discrete reflections as well as to reduce the general reverberation time. The result was a reduction in the measured time to 3.5 seconds and the complete elimination of acoustic difficulties. Present reverberation formulas do not possess sufficient generality to justify application to enclosures which are extremely atypical in size or shape. Until such formulas are developed, reliance must be placed on actual measurements.

---

[5] *Electrical Engineering*, March, 1932.
[6] *Proc. I. R. E.*, April, 1932.
[7] *Jour. S. M. P. E.*, May, 1932.
[8] *Jour. S. M. P. E.*, April, 1932.

Contributors to this Issue

C. B. H. Feldman, B.Sc., University of Minnesota, 1926; Teaching Fellow, University of Minnesota, 1926–28; M.Sc., University of Minnesota, 1928. Bell Telephone Laboratories, 1928–. Mr. Feldman has been engaged in short-wave radio receiving. His work has been mainly on transmission lines, antennas, and wave propagation problems.

P. B. Flanders, A.B., Harvard University, 1920; S.B. in Electrical Engineering, 1922. Engineering Department, Western Electric Company, 1922–25; Bell Telephone Laboratories, 1925–. Mr. Flanders has worked principally on acoustic devices and on the development of acoustic measuring methods.

Bancroft Gherardi, B.Sc., Polytechnic Institute, Brooklyn, N. Y., 1891; M.E., Cornell University, 1893; M.M.E., Cornell University, 1894. New York Telephone Company, Engineering Assistant, 1895–99; Traffic Engineer, 1899–1900. New York and New Jersey Telephone Company, Chief Engineer, 1900–06. New York Telephone Company, and New York and New Jersey Telephone Company, Assistant Chief Engineer, 1906–07. American Telephone and Telegraph Company, Equipment Engineer, 1907–09; Engineer of Plant, 1909–18; Acting Chief Engineer, 1918–19; Chief Engineer, 1919–20; Vice President and Chief Engineer, 1920–. Mr. Gherardi is a Past President of the American Institute of Electrical Engineers and is now President of the American Standards Association.

H. C. Harrison, A.B., Colorado College, 1910; S.B., Massachusetts Institute of Technology, 1913; Instructor in Electrical Engineering, Massachusetts Institute of Technology, 1913–14. Western Electric Company, 1914–24; Bell Telephone Laboratories, 1925–. As Special Transmission Instruments Research Engineer, Mr. Harrison has been instrumental in applying the principles of electric transmission theory to sound recording and reproducing equipment.

M. J. Kelly, B.S., Missouri University, 1914; M.S., University of Kentucky, 1915; Ph.D., University of Chicago, 1919; Instructor of Physics, University of Kentucky, 1914–15; Research Assistant, University of Chicago, 1915–18. Western Electric Company, 1918–25; Bell Telephone Laboratories, 1925–. As Vacuum Tube Development

Engineer, Dr. Kelly has worked on the development of thermionic and photoelectric devices.

D. Mitchell, B.S., Princeton University, 1925. Department of Development and Research, American Telephone and Telegraph Company, 1925–. Mr. Mitchell has been engaged in transmission development work on voice-operated devices and problems of interconnecting wire and radio telephone systems.

H. H. Nance, Washington University, 1906–10. American Telephone and Telegraph Company, Long Lines Plant Department, 1910–23; Long Lines Engineering Department, Acting Engineer of Transmission, 1923; Engineer of Transmission, 1924–1928; Plant Extension Engineer, 1928–.

C. H. Prescott, Jr., B.A., Yale University, 1922; Ph.D., California Institute of Technology, 1926; Research Fellow, California Institute of Technology, 1926–27; National Research Fellow in Chemistry, 1927–28. Bell Telephone Laboratories, 1928–. Dr. Prescott has been engaged in development and research on photoelectric and thermionic vacuum tubes.

K. L. Scott, B.S. in Electrical Engineering, University of Wisconsin, 1921; M.S. in E.E., 1928; Instructor in Electrical Engineering, University of Wisconsin, 1921–24. Engineer, Western Electric Company, 1926–. Mr. Scott has been engaged in laboratory investigations of magnet steels and permanent magnets, and the development of manufacturing processes and equipment in connection with magnet manufacture.

E. J. Sterba, B.E. in Electrical Engineering, State University of Iowa, 1920. Engineering Department, Western Electric Company, 1920–25; Bell Telephone Laboratories, 1925–. Mr. Sterba has been engaged chiefly in the development of transmitting antennas and apparatus related to antennas.

S. B. Wright, M.E., M.E.E., Cornell University, 1919. Department of Development and Research, American Telephone and Telegraph Company, 1919–. Mr. Wright has been engaged in transmission development work on voice-operated devices and problems of interconnecting wire and radio telephone systems.

The Bell System Technical Journal

October, 1932

World-Wide Telephony—Its Problems and Future *

By BANCROFT GHERARDI and FRANK B. JEWETT

The rapid development of large telephone networks giving a high grade of service between large numbers of telephones in continental areas has laid the foundation for the development of world-wide telephone service. Beginning in 1927 with the establishment of the first commercial telephone circuit between Europe and North America, intercontinental telephone service has, during the past five years, extended rapidly, and with further extensions already definitely planned, will embrace all of the continents and make possible the connection together of practically any two telephones in the world.

Up to the present time radio has been used to overcome the peculiar technical obstacles in the provision of intercontinental circuits. Two portions of the radio spectrum are suitable for this purpose, the long wave providing only a few circuits and the short wave providing for possibly several hundred circuits in the world as a whole. Plans have already been made for the important route between Europe and North America to supplement these with a telephone cable and the use of wire lines for intercontinental routes may become more important in the future.

The full development of intercontinental telephony is affected by a number of general difficulties. Of these the differences in time between different parts of the earth's surface are inherent. Differences in language both affect the ease with which customers can converse over the telephone, and complicate the operating problem. Furthermore, some of the differences in operating and commercial practices in the telephone networks of different continents which have in the past developed largely independently of each other, require consideration in the building up of intercontinental services.

The full development of intercontinental telephony is dependent upon the continued progress in working out these problems and in an extension of the brilliant scientific and engineering achievements which have made possible the present services. It is to be expected that with the further growth of intercontinental service it will be found desirable in the future to adopt a general world-wide plan for the routing of intercontinental messages somewhat comparable to the plans for continental telephone service already under consideration or in use. While political considerations may temporarily affect the form of the world-wide network, ultimately the requirements of economy and good service will no doubt be determining factors in such a plan.

It is to be hoped that the continued closer knitting together of the nations and races of the world by intercontinental telephone circuits will be a great contribution to international friendship and good will.

The authors wish to acknowledge their indebtedness to a number of telephone administrations who have provided them with information regarding present and proposed intercontinental services, supplementing the data previously published. They have also drawn freely on the material presented in the bibliography of Appendix 2, and this material has been of assistance. The authors also express appreciation of the assistance given them by a number of their associates in the American Telephone and Telegraph Company, particularly Messrs. O. B. Blackwell, A. B. Clark, L. Espenschied, O. T. Laube, H. S. Osborne and H. E. Shreeve.

---

* Presented at The International Electrical Congress, Paris, France, July 5–12, 1932.

INTRODUCTION

THE rapidity with which telephone service has been extended to world-wide proportions during the past few years is perhaps one of the most remarkable of man's conquests over time and distance. Already it is a commonplace to hear the human voice from thousands of miles away, over land and sea. Distance and the great natural barriers of the world no longer prevent us from talking with each other. Across oceans and over high sierras the voice now carries its full message. Furthermore, these results have been accomplished within a few years. Telephony has demonstrated its international and intercontinental services. The development and extension of those services lies before us. It seems an appropriate time, therefore, to review some of the problems and possibilities in this extension of the application of the electrical arts to the service of mankind.

World-wide telephony has as its foundation the wire line networks on the various continents with their millions of users. The first step was taken toward overcoming the great barrier presented by the oceans in 1891 when the first submarine telephone cable was laid between Dover and Calais. Further submarine cable developments followed including the laying of a continuously loaded cable between Denmark and Sweden in 1902.

In the meantime the problems of spanning great distances over land were being rapidly solved, both in North America and in Europe, through the application of the loading coil and other transmission improvements. In 1911 service was opened between New York and Denver, a distance of 2000 miles (3300 kilometers). By 1913 the development of underground cable had progressed so that a cable was placed in service between Boston, New York and Washington, a distance of over 420 miles (700 kilometers).

But it required still further improvements in the whole art of long distance telephone transmission, including amplifiers and their application to wire lines before long distance telephony could develop beyond the semi-continental stage to truly transcontinental and transoceanic distances. With improved amplifier elements and the perfection of means for applying repeaters came the opening of transcontinental service in America in 1915 initially between New York and San Francisco, 3200 miles (5300 kilometers). This was followed after the close of the war by the rapid development of telephony of continental scope throughout Europe, stimulated by the close cooperation of the European Administrations through the International Advisory Committee on Long Distance Telephony. In South America the trans-Andean telephone line between Buenos Aires and Santiago was opened in 1928. Two years later came the transcontinental telephone line in Australia.

Overseas radio telephone experiments, in 1915, successfully transmitted the human voice from Washington to Paris and from Washington to the Hawaiian Islands. Commercial development of intercontinental telephony, however, followed somewhat slowly both because of the war and because of the tremendous inherent technical difficulties. However, as an interesting fore-runner of what would come later, a public service radiotelephone system was opened in 1920 linking Catalina Island, off the coast of California, with the wire telephone network of North America. In 1927, after a period of intensive experimentation and development work, commercial service between Europe and North America was opened to the public. This was the first step in the expansion of telephone service from a continental scope to a world-wide scope.

Present Situation and Plans for Future Development

Intercontinental telephony has naturally been dependent upon and been preceded by the development of large networks of telephones to which the intercontinental circuits could be connected. The most highly industrialized parts of the earth's surface today are provided with extensive networks of telephone lines covering great areas and interconnecting large numbers of telephones. This is indicated in Fig. 1 which shows the principal national and continental wire telephone networks, and all individual cities not connected to an extensive interurban network which according to latest reports have more than 10,000 telephones. As a result of the improvements which have been made in these wire networks in the last 15 years, both as to transmission and speed of service, they are today generally available and satisfactory for use in connection with intercontinental service.

For the purpose of this paper circuits of less than 600 miles (1000 kilometers) even if they cross continental boundaries will not be considered intercontinental circuits. In the consideration of world-wide telephony we are interested in the long circuits between distant parts of different continents or between continental areas and distant islands, for which the technique developed for long continental telephone lines is not directly applicable.

The extent to which the continents of the world are already interconnected by telephone circuits and the additional connections contemplated by plans now under way are indicated in Fig. 2. These are such as to make possible conversations between any two continents of the earth either by direct circuit or by switching in Europe or in

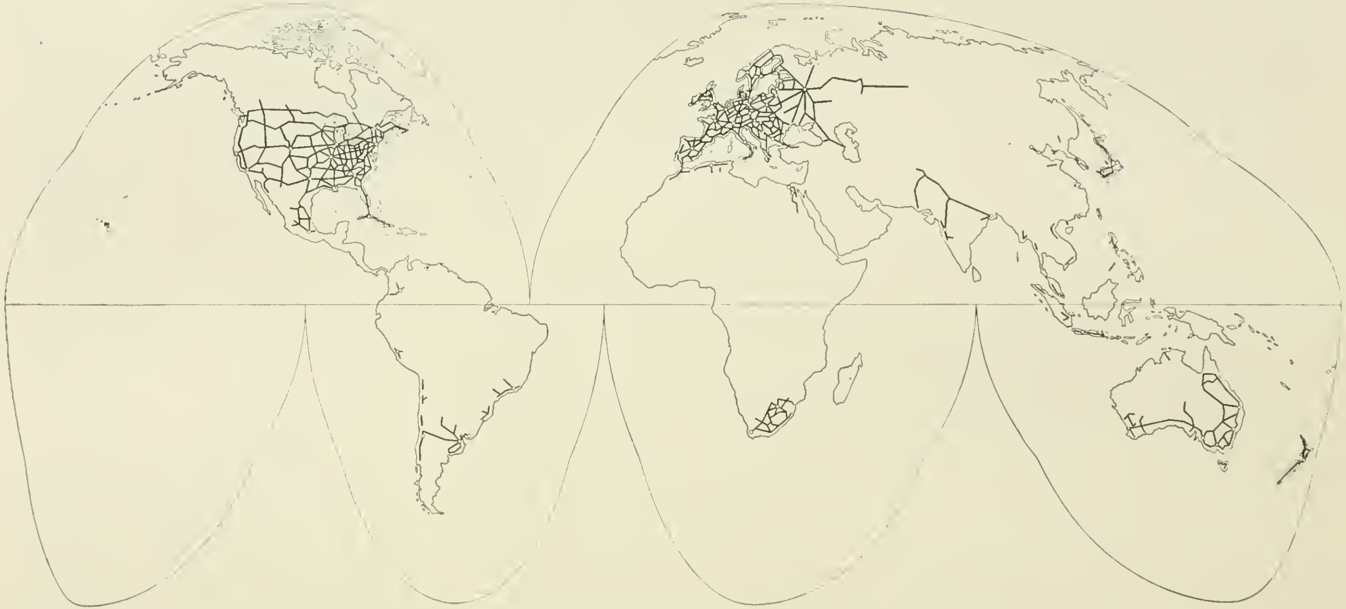
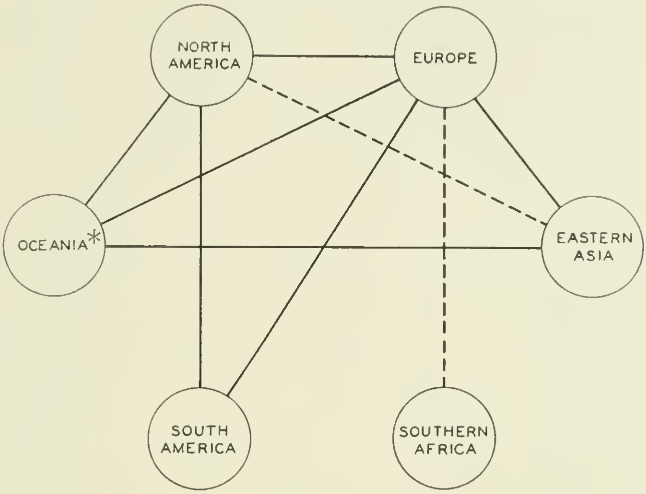


Fig. 1—Wire line telephone networks of the world. Includes indication of areas covered by international, national or interurban networks and individual cities having over 10,000 telephones. Based on information available January 1, 1932.

North America. The existing intercontinental telephone circuits of the world are indicated in Fig. 3. In this figure distinctions are made between full time circuits and part time circuits, that is, those on which terminal apparatus is shared by different points. A distinction is also made between circuits which interconnect wire networks connecting with 20,000 telephones or more, and those terminating in single telephones or networks of less than 20,000 telephones.



*THE EXISTING DIRECT CONNECTIONS ARE AS FOLLOWS:
EUROPE TO AUSTRALIA AND JAVA
NORTH AMERICA TO HAWAII
EASTERN ASIA TO JAVA

AT PRESENT CONNECTIONS BETWEEN THE AMERICAS AND AUSTRALIA AND JAVA ARE ESTABLISHED BY WAY OF EUROPE

——— EXISTING DIRECT CONNECTIONS
- - - - - PROPOSED DIRECT CONNECTIONS

Fig. 2—International telephone relations as of January 1, 1932.

There are at present 37 of these intercontinental circuits totalling about 168,000 miles (280,000 kilometers) in length. All are radio circuits. One, in the New York-London group, is a long-wave circuit operating at approximately 60 kilocycles; the others are short-wave circuits in the range between 6000 and 23,000 kilocycles.

Europe and North America, which are the two largest highly industrialized areas of the world, contain about 90 per cent. of the world's telephones. It is natural that intercontinental telephone business in volume should first develop between those two areas Here service is maintained on a 24-hour basis and a group of four circuits is in use. Elsewhere, however, intercontinental connections consist at the present time of a single circuit, or more frequently,

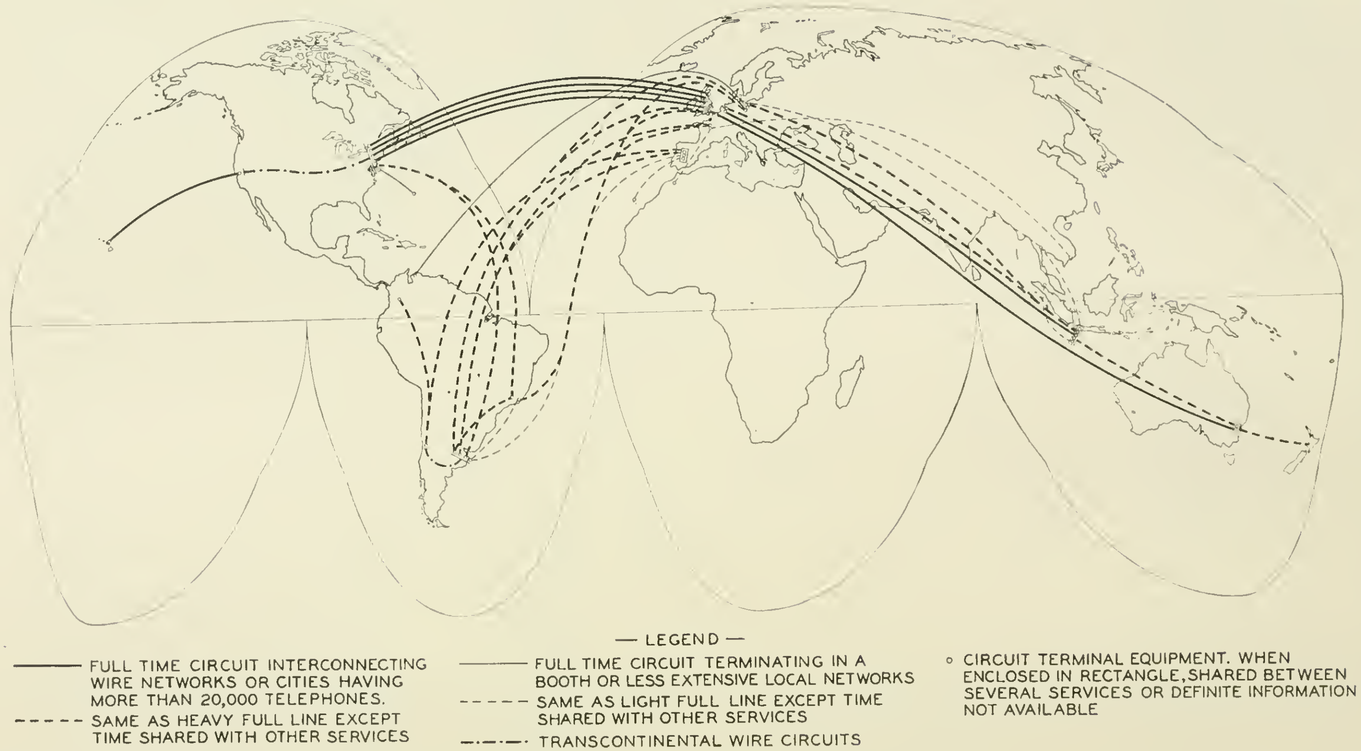

— LEGEND —
FULL TIME CIRCUIT INTERCONNECTING WIRE NETWORKS OR CITIES HAVING MORE THAN 20,000 TELEPHONES.
SAME AS HEAVY FULL LINE EXCEPT TIME SHARED WITH OTHER SERVICES
FULL TIME CIRCUIT TERMINATING IN A BOOTH OR LESS EXTENSIVE LOCAL NETWORKS
SAME AS LIGHT FULL LINE EXCEPT TIME SHARED WITH OTHER SERVICES
TRANSCONTINENTAL WIRE CIRCUITS
○ CIRCUIT TERMINAL EQUIPMENT. WHEN ENCLOSED IN RECTANGLE, SHARED BETWEEN SEVERAL SERVICES OR DEFINITE INFORMATION NOT AVAILABLE

Fig. 3—Long intercontinental telephone circuits of the world. Circuits over 600 miles (1000 kilometers) in length as of January 1, 1932.

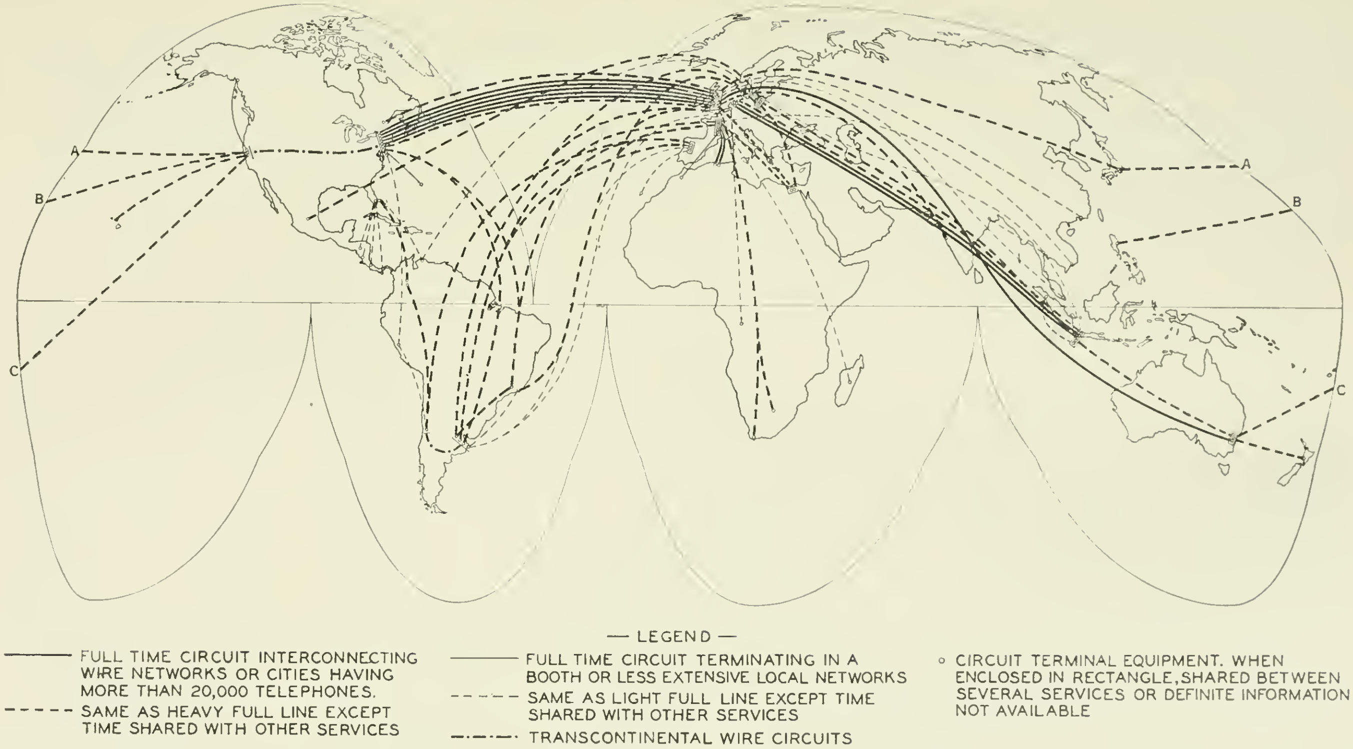

A
B
C
A
B
C
— LEGEND —
FULL TIME CIRCUIT INTERCONNECTING WIRE NETWORKS OR CITIES HAVING MORE THAN 20,000 TELEPHONES.
SAME AS HEAVY FULL LINE EXCEPT TIME SHARED WITH OTHER SERVICES
FULL TIME CIRCUIT TERMINATING IN A BOOTH OR LESS EXTENSIVE LOCAL NETWORKS
SAME AS LIGHT FULL LINE EXCEPT TIME SHARED WITH OTHER SERVICES
TRANSCONTINENTAL WIRE CIRCUITS
CIRCUIT TERMINAL EQUIPMENT. WHEN ENCLOSED IN RECTANGLE, SHARED BETWEEN SEVERAL SERVICES OR DEFINITE INFORMATION NOT AVAILABLE

Fig. 4—Existing and proposed long intercontinental telephone circuits of the world. This map is similar to Fig. 3 and includes existing circuits and additional circuits now definitely proposed. Based on information available January 1, 1932.

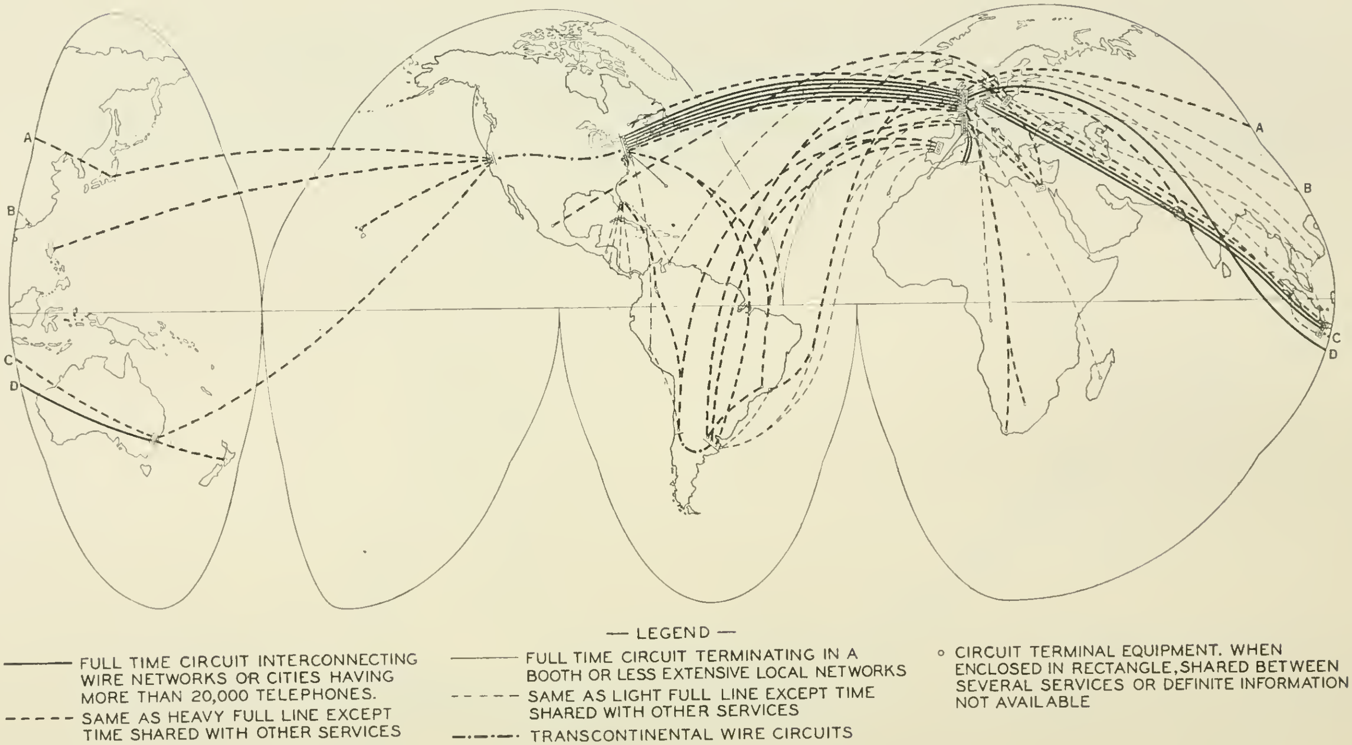

A
B
C
D
A
B
C
D
— LEGEND —
FULL TIME CIRCUIT INTERCONNECTING WIRE NETWORKS OR CITIES HAVING MORE THAN 20,000 TELEPHONES.
SAME AS HEAVY FULL LINE EXCEPT TIME SHARED WITH OTHER SERVICES
FULL TIME CIRCUIT TERMINATING IN A BOOTH OR LESS EXTENSIVE LOCAL NETWORKS
SAME AS LIGHT FULL LINE EXCEPT TIME SHARED WITH OTHER SERVICES
TRANSCONTINENTAL WIRE CIRCUITS
○ CIRCUIT TERMINAL EQUIPMENT. WHEN ENCLOSED IN RECTANGLE, SHARED BETWEEN SEVERAL SERVICES OR DEFINITE INFORMATION NOT AVAILABLE

Fig. 5—Existing and proposed long intercontinental telephone circuits of the world. Arranged to show transpacific connections. Based on information available January 1, 1932. Circuits and legend same as shown in Fig. 4.

the part time use of a circuit, the apparatus being arranged to share its time between two or more distant terminals.

One cannot help being impressed by the developments represented by Fig. 3, which have taken place in the short space of five years. We are, nevertheless, in an early stage of development and present service suffers from the limitations of pioneer conditions. Dependence in so many cases upon a single intercontinental circuit which often is shared on a "party line" basis between two or more terminal points, is one of these limitations. Another is the variation in transmission characteristics and the susceptibility to interruptions which in the present stage of development of radio are characteristic especially of short-wave circuits on some of the more important routes. The rapidity with which the first steps have been taken is perhaps the best assurance that advance will continue to be rapid as the commercial demand for increasing amounts of service develops. This is illustrated by Figs. 4 and 5 which show the intercontinental circuits of the world which will exist when the present plans for additional circuits have been completed in so far as these plans are known to the authors.

Fig. 6 shows the extent to which the countries of the world will be tied into one great telephone system when the plans outlined in Figs. 4 and 5 are completed. Fig. 6 also shows the distribution of world trade and population of various groups of countries which are so interconnected. It will be noticed that the world-wide telephone network will include countries having 99 per cent of all the telephones of the world and having 92 per cent of the world's foreign trade. Details of all the present and proposed circuits which have come to the knowledge of the authors are given in Appendix I.

Closely related to the establishment of intercontinental circuits is the establishment of telephone service from European and North American points to a number of passenger vessels normally operating on the North Atlantic and other passenger routes. In addition, work is advancing in the equipment of fishing fleets, tugs, etc., but this service pertains more directly to the continental telephone service.

Technical Problems and Limitations

The unusual technical difficulties of providing satisfactory intercontinental telephone circuits come about not merely because of the distances involved but because these distances include long stretches of sea or undeveloped land. For years prior to the establishment of the first intercontinental circuit, commercial service was given on continental telephone networks over comparable distances, for example, in North America up to about 6000 miles (10,000 kilometers).

WORLD DISTRIBUTION

POPULATION, TELEPHONES, FOREIGN TRADE

For Countries Having Existing or Proposed International Telephone Connections as of January 1, 1932

Country	Total Foreign Trade 1930 (Millions of Dollars)	Total Telephones as of Jan. 1, 1931 (Thousands)	Total Population (Thousands)
EUROPE			
All Principal Countries	30,260	10,620	542,350
NORTH AMERICA			
United States, Canada, Cuba, Mexico, Bermuda	9,500	21,768	153,780
SOUTH AMERICA			
Argentina, Brazil, Chile, Colombia, Uruguay, Venezuela	2,590	595	71,090
OCEANIA			
Hawaii, Australia, New Zealand, Java, Sumatra	2,310	760	69,440
ASIA			
Siam, Indo-China	300	12	31,740
AFRICA			
Morocco, Canary Islands	120	14	5,850
Total in countries having existing connections to international networks	45,080	33,769	874,250
NORTH AMERICA			
Central America and Bahamas	120	18	2,720
SOUTH AMERICA			
Peru	130	14	6,260
OCEANIA			
Philippines	260	26	12,700
ASIA			
Japan, Hong Kong, British India, Malay States	3,810	1,028	410,320
AFRICA			
Algeria, Belgian Congo, Egypt, Union of So. Africa, Madagascar	1,370	197	47,700
Total in countries having prospective international connections	5,690	1,283	479,700
Total, present and prospective international network	50,770	35,052	1,353,950
All Other	4,140	308	622,550
Total World	54,910	35,360	1,976,500

Fig. 6.

The technique of transmission over long continental distances, however, includes as one fundamental element the use of intermediate amplification at frequent intervals of from 45 to 240 miles (75 to 400 kilometers), depending upon the electrical characteristics of the circuit.

Radio Links

All of the existing intercontinental circuits make use of radio for the long transoceanic jumps, and it seems desirable first to consider the technical problems of radio telephony.

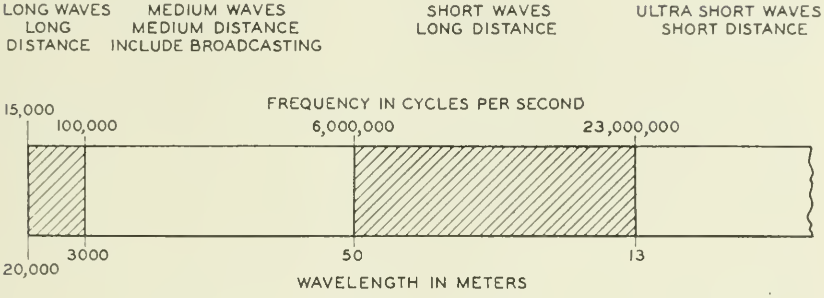


Fig. 7—Radio frequency chart. The bands applicable to intercontinental service are indicated by shaded areas.

Fig. 7 shows the radio spectrum with an indication of the uses made of the various portions of the spectrum. Only the two shaded parts of this spectrum appear to be applicable to intercontinental telephony. One, in the long wave range, extends from about 15 kilocycles to perhaps 100 kilocycles. The second includes the short-wave range from about 6000 to 23,000 kilocycles.

The transmission of these two wave ranges exhibit interesting differences in characteristics. For the short wave-lengths the transmission is frequently referred to as being in the form of "sky waves." This is for the reason that at intermediate distances the waves may practically disappear near the surface of the ground but reappear at greater distances. They appear to have been carried around the curvature of the earth's surface by refraction or reflection from ionized atmospheric layers. While the action in the long-wave range at great distances appears to be also conditioned partly by the ionized layers, the field at the surface of the earth falls off continuously as the distance from the transmitter is increased.

Interesting results are now being obtained in the use for communication purposes of very short radio waves having frequencies above 30,000 kilocycles. The work done to date indicates that these frequencies are not sufficiently deflected from their paths by the atmosphere to follow the earth's curvature. This characteristic appears to prevent the use of these very short radio waves for direct transmission over long distances and limits direct transmission to distances so short that the earth's curvature is not a large factor. For this reason the very high frequency range is not indicated in Fig. 7 as suitable for intercontinental circuits. Where the route is over-land it is possible that such rays may find practical use in forming links in intercontinental circuits particularly where the topography of the country affords advantageous elevated locations for intermediate repeater points. Too little is now known of these very short waves to make their discussion other than speculative.

An important requirement for radio for overseas telephone circuits is the avoidance of overhearing. A number of the overseas circuits now in operation, including the transatlantic group, are equipped with privacy arrangements which so modify the frequency disposition of the voice waves as to prevent overhearing of the conversations by other radio stations not equipped with similar arrangements. Experiments have been made with more elaborate arrangements for obtaining an even higher degree of privacy than that now provided.

Long Wave Circuits

Experience has shown that good results can be obtained in the long-wave range on such circuits as those between North America and Europe which are 3000 miles (5000 kilometers) long and have their transmission paths at a high latitude. A fundamental limitation to the extensive use of long waves, however, is that within the range of suitable frequencies there are theoretically obtainable only about twenty telephone channels between any two points. Actually however, there are not this many available because of the practical limitations imposed by the necessity of sharing this range of frequencies with other types of radio service. The relatively low attenuation of these waves as they are transmitted over the earth's surface makes it appear impractical to use duplicate intercontinental circuits of the same frequency at different points on the earth's surface. The total number of circuits of the long-wave type which can form a part of the ultimate world-wide telephone network seems, therefore, to be very small.

Other limitations in the use of long-wave circuits are that with the amounts of power which it now seems practicable to use (a tube capacity of about 300 kilowatts is used in the transmitter of the present transatlantic long wave circuit) their successful application does not exceed several thousand kilometers and is confined to routes well outside of the equatorial regions. These limitations are necessary to avoid excessive interference from atmospheric disturbances, both because of the relatively large components of these disturbances having frequencies in the range of wave-lengths used by these circuits, and also because of the difficulty of obtaining a high degree of directivity. Advantage has been obtained in the transatlantic long-wave channel by locating directive receiving antennas in the most northerly portions of the United States and Great Britain.

Short-Wave Radio

The short-wave range, as indicated in Fig. 7, covers a very much broader band of frequencies and offers hope for a much larger number of circuits. Only a portion of the entire range can be used at any one time for circuits between any two points because the distance at which the refracted waves reach the earth's surface varies with the frequency. It also varies with the time of day and the season of year, making it necessary, in general, to have about three different wave-lengths allocated for each circuit. In view of this limitation, it appears that this range can provide theoretically perhaps 50 telephone circuits between two given points. This theoretical possibility is reduced by practical considerations and by the necessity of sharing the range with other radio services.

Due, however, to the fact that short waves are more restricted than long waves in distance and in the time of day for which they are effective, it may be possible to use the same frequency for different circuits in several parts of the world without interference. It seems now that the development of the present art may make available throughout the entire world short-wave intercontinental circuits numbered in the hundreds but not in the thousands. This is not an inexhaustible supply of channels for the world to share as is illustrated by the very much larger number of telephone circuits already required for long distance service in the great continental telephone networks. In the United States alone there are about 6500 circuits in long distance service of which 620 are over 600 miles (1000 kilometers) in length.

The fact that by proper selection of frequency it is possible to communicate over the longest distances, gives short waves a flexibility which is unique. They give particularly good results in routes crossing the equator, such as circuits between Europe or North America and South America, for which routes long-wave transmission would be impracticable because of heavy atmospheric disturbances. On the other hand, short-wave transmission is subject to variations and interruptions from fading, distortion, and effects connected with terrestrial magnetic disturbances. These effects appear to be particularly severe on circuits involving routes of transit near the polar regions, such as the route between North America and Europe. Some of these effects are illustrated in Figs. 8 and 9 which give data taken on the United States-Europe route and which show, for comparison, the relatively stable conditions for the long-wave channel. In many respects short-wave and long-wave radio channels admirably supplement each other, being seldom interrupted at the same time.

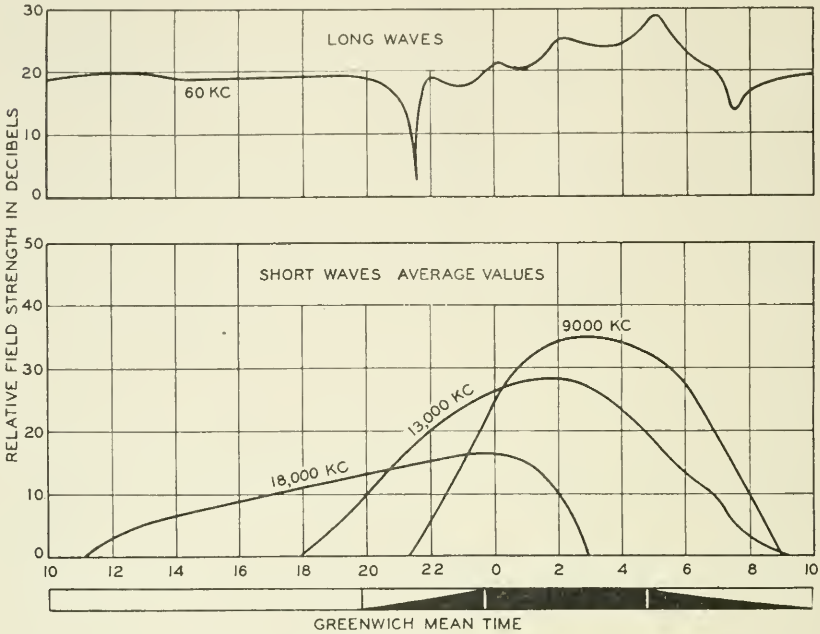


Fig. 8—Diurnal field strength characteristics of long and short-wave transmission over the North Atlantic. Shows effect of sunrise and sunset on the long wave and the necessity with short waves for using different wave-lengths at different times of day.

Short-wave transmission is in an early stage of development and it may be hoped that, with further experience, means will be found to avoid or reduce at least some of the present major limitations.

Wire Circuits

The intercontinental routes offer great difficulties to the placing of wire circuits. However, the limitations of radio, particularly on the intercontinental routes between North America and Europe interconnecting the two largest continental groups of telephones, have naturally led to considerable engineering thought on the possibilities of wire circuits. The most interesting development in this connection is a submarine cable which can be used for long lengths without intermediate repeaters.

Long Submarine Cables Without Intermediate Repeaters

To design long submarine cables without repeaters requires providing the mechanical characteristics necessary for deep sea cables and meeting rather definite limits of overall attenuation. There is a limit to the power which can be imposed on the sending end of the cable to avoid excessive voltage stress on the cable insulation. Even though the cable be adequately shielded from extraneous noises, there is a minimum amount of received power with which satisfactory transmission can be obtained because of the thermal agitation in the conductor itself. These two limits result in a maximum attenuation of about 160 db for satisfactory operation over such a cable.

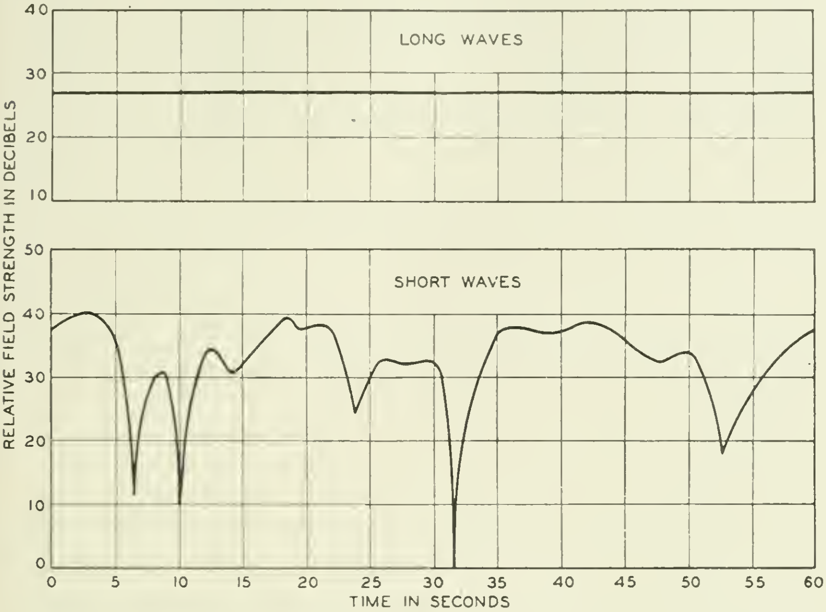


Fig. 9—The upper curve indicates the short period stability of the long waves. The lower curve indicates the rapid variations which are frequently experienced with short-wave transmission.

The actual design of intercontinental telephone cables has been largely directed toward the North Atlantic route between the United States and Europe, including a direct section from Newfoundland to Ireland, a distance of about 2100 miles (3400 kilometers). By the use of a new insulating material, paragutta, with dielectric losses about one-thirtieth those of guttapercha, but with similar mechanical characteristics, and by the use of perminvar, a new magnetic material for loading the cable, it has been possible to design a cable for this section meeting both the mechanical and electrical requirements. With the further development of intercontinental communications, it is proposed to lay such a cable within a few years. With the materials now available, such a cable will provide only one channel of communication, and two-way communication will be carried out by using it as a one-way channel automatically switched in the direction of transmission by the voices of the users. Transatlantic cables providing several telephone channels either by the use of separate conductors or by carrier current methods are beyond the range of the present art.

The direct submarine cable while relatively expensive, is expected to provide a high grade circuit having a stability and a freedom from interruption greater than that provided by the present radio circuits. Proper combination of three different types of circuit, long and short-wave radio and cable, in one circuit group, however, should provide a large measure of assurance of continuous high grade service not dependent on the troubles which may affect any one type of circuit, and at an average expense not greatly above that required for radio circuits alone.

Other Wire Routes

The technical possibilities of wire circuits on intercontinental routes are evidently greatly increased where means can be found to avoid long lengths of submarine cables by the use of intermediate repeater stations. On the direct route between Europe and America, nature has not been kind enough to supply a series of islands at convenient distances and reasonably low latitudes to serve as repeater points. By going north, long lengths of submarine could be avoided. For example, a route through Greenland, Iceland and the Faroe Islands could be laid out with a maximum length of submarine cable of 300 miles (500 kilometers). The obvious difficulties of the route are the great extents of inaccessible and sparsely settled country, and the placing and maintenance of submarine cable under very difficult fog, storm and ice conditions. In the early days of submarine telegraphy such a northern route was seriously considered before the cable art had reached the point of permitting direct cables, but it was never used.

An equally bold solution which has been proposed is to float the desired repeater stations in the open sea. A good deal of ingenuity has been exercised in considering the possibilities of both attended and unattended floating repeater stations and of stations submerged below the action of the waves. So far, however, it is by no means clear that the mechanical difficulties and the problems of maintenance can be dealt with satisfactorily.

As time goes on land lines may become possible for many intercontinental routes where now, because of lack of highways or railways and the wild and unsettled nature of the country, they are out of the question.

OPERATING AND COMMERCIAL PROBLEMS

In addition to the technical problems and limitations associated with the development of a world-wide telephone service which are outlined in the foregoing section, there are numerous operating and commercial problems and still other difficulties of a general nature.

Differences in Time

Limitations of a fundamental character are imposed by the differences in time on different parts of the earth's surface. These differences for a number of the principal metropolitan centers of the world are indicated in Fig. 10. If the business day be assumed to be 8 hours long, it is evident that as a result of the differences in time there is for each city one-third of the earth's surface on which the time is so different from that of the city in question that there is no overlap of the business day. For western Europe this third of the world is for the most part in the Pacific Ocean, so that, as will be noted from Fig. 10, there are few important centers in the world in which the time is more than 8 hours different from that of western Europe. The western part of the United States, however, has time differences of more than 8 hours with a large part of Europe, Asia and Africa. During some portion of the waking day there is an overlap of time between any two points on the earth.

Language

Another limitation, though of a less inherent nature, is the difference in language between different nations. There is as yet no evidence that the electrical requirements for satisfactory transmission of speech are substantially different with different languages, but language differences produce a problem in many intercontinental telephone conversations as often one subscriber is using a language of which he is not wholly master, and this may sometimes apply to both. Under these conditions the transmission requirements of the circuit for a given ease of carrying on conversations is unquestionably greater than when the subscribers are conversing in their native tongue. Hence, there is need in the ultimate development of intercontinental telephone systems for a high standard of transmission.

Language differences also complicate the operating problem, as it is important that the operators understand each other easily. This is

Hours Time Difference Between Certain Cities
in Various Parts of the World

A → B ↓	Los Angeles San Francisco Vancouver	Chicago New Orleans Mexico City * Winnipeg	New York Washington Montreal Santiago	Buenos Aires Halifax	Rio de Janeiro	Paris London Madrid
Los Angeles San Francisco Vancouver	–	+ 2	+ 3	+ 4	+ 5	+ 8
Chicago New Orleans Mexico City * Winnipeg	– 2	–	+ 1	+ 2	+ 3	+ 6
New York Washington Montreal Santiago	– 3	– 1	–	+ 1	+ 2	+ 5
Buenos Aires Halifax	– 4	– 2	– 1	–	+ 1	+ 4
Rio de Janeiro	– 5	– 3	– 2	– 1	–	+ 3
Paris London Madrid	– 8	– 6	– 5	– 4	– 3	–
Berlin Rome Stockholm Vienna	– 9	– 7	– 6	– 5	– 4	– 1
Cairo Capetown Istambul Moscow	–10	– 8	– 7	– 6	– 5	– 2
Bombay Delhi	+10.5	–11.5	–10.5	– 9.5	– 8.5	– 5.5
Bangkok Singapore	+ 9	+11	+12	–11	–10	– 7
Peiping Shanghai Manila Perth	+ 8	+10	+11	+12	–11	– 8
Tokyo Vladivostock	+ 7	+ 9	+10	+11	+12	– 9
Sydney Melbourne	+ 6	+ 8	+ 9	+10	+11	–10

Notes

Computed from "The Time Zone Chart of the World," published by the Navy Department of the United States.

+ Indicates that cities in the row designated "A" have later clock time than those in the column designated "B" and lie within 180° toward the east.

Fig. 10

A → B ↓	Berlin Rome Stockholm Vienna	Cairo Capetown Istambul Moscow	Bombay Delhi	Bangkok Singapore	Peiping Shanghai Manila Perth	Tokyo Vladivostock	Sydney Melbourne
Los Angeles San Francisco Vancouver	+ 9	+10	−10.5	− 9	− 8	− 7	− 6
Chicago New Orleans Mexico City Winnipeg	+ 7	+ 8	+11.5	−11	−10	− 9	− 8
New York Washington Montreal Santiago	+ 6	+ 7	+10.5	−12	−11	−10	− 9
Buenos Aires Halifax	+ 5	+ 6	+ 9.5	+11	−12	−11	−10
Rio de Janeiro	+ 4	+ 5	+ 8.5	+10	+11	−12	−11
Paris London Madrid	+ 1	+ 2	+ 5.5	+ 7	+ 8	+ 9	+10
Berlin Rome Stockholm Vienna	−	+ 1	+ 4.5	+ 6	+ 7	+ 8	+ 9
Cairo Capetown Istambul Moscow	− 1	−	+ 3.5	+ 5	+ 6	+ 7	+ 8
Bombay Delhi	− 4.5	− 3.5	−	+ 1.5	+ 2.5	+ 3.5	+ 4.5
Bangkok Singapore	− 6	− 5	− 1.5	−	+ 1	+ 2	+ 3
Peiping Shanghai Manila Perth	− 7	− 6	− 2.5	− 1	−	+ 1	+ 2
Tokyo Vladivostock	− 8	− 7	− 3.5	− 2	− 1	−	+ 1
Sydney Melbourne	− 9	− 8	− 4.5	− 3	− 2	− 1	−

— Indicates that cities in the row designated "A" have earlier clock time than those in the column designated "B" and lie within 180° toward the west.

This does not take account of the change in date in crossing the International date line.

* Time after April 1, 1932.

Fig. 10 (continued).

accomplished to a large extent at the present time by providing bi-lingual operators at the terminals of international circuits. Even so, on built-up connections with switches in several countries, operation is cumbersome since the terminal operators may not be able to talk to each other directly. This is one of the problems to which the International Advisory Committee on Long Distance Telephony has been giving attention. It is out of the question in a world-wide telephone system to expect the operator handling the call to be able in all cases to talk directly with the distant subscriber. However, it seems important from the standpoint of giving the highest grade of service that ultimately she should be able in all cases to talk directly and easily with the operator at the distant national terminal, no matter where this may be.

Operating and Commercial Practices

The continental telephone networks have in the past developed to a large extent independently of each other, and it is therefore natural that there should now exist differences in operating and commercial practices, some of which must be considered in giving intercontinental service. In Europe and in the United States, to take, for example, the two largest networks, the point of view in the development of the telephone systems has been somewhat different. In Europe emphasis has been laid on the continuous use of the long toll circuits, developing for them as large a message capacity as possible, while in the United States emphasis has been given to the rapid completion of all calls. This difference in point of view naturally led to differences in practice. For example, the classification of service based upon urgency and the limitation in length of conversations generally in use in Europe are not found in American practice. Also, an important factor in American practice for connections over long distances is the so-called person-to-person service in which a specified person is called rather than a specified telephone number, and the order is considered satisfied only when the person called is brought to the telephone. This service has not been generally used in European practice. These serve as illustrations of the type of difference in practice which must be adjusted between the administrations involved in establishing new intercontinental services, and it is evident that as these services become more used and more nearly universal, these adjustments will become increasingly important.

A consideration which might be important in the development of intercontinental services in some cases is a difference in standards of transmission or of speed of completion of calls. In recent years there has been in all large telephone networks a trend towards higher standards of service and this has been a favorable factor in making possible the beginnings of an intercontinental service.

GENERAL DISCUSSION AND CONCLUSIONS

The extension of overseas telephony during the past five years has already linked together into one system all of the largest continental telephone networks of the world and with the completion of further extensions now under way, this world-wide system will include all but two of the wire networks which give access to more than 20,000 telephones in all six continents of the earth. True, many of these overseas connections are as yet but slender threads of conversation, important perhaps not so much because of the communication which they now handle but because they represent the first realization of great possibilities in the achievement of a world-wide telephone system closely linking together the continents of the earth. In the words of Mr. Walter S. Gifford, President of the American Telephone and Telegraph Company, in the Annual Report of that company for the year 1926, the ultimate ideal of a world-wide system is that it shall enable "anyone anywhere to pick up a telephone and talk to anyone else anywhere else, clearly, quickly and at a reasonable cost."

What are the obstacles to the realization of this ideal for world-wide telephony? Intercontinental service is subject to extraordinary technical and operating difficulties and as yet only the first steps in overcoming these difficulties have been taken. The quality of service at the present time both as regards transmission, continuity and speed of service is not comparable with that given today on the large continental telephone networks, but is more comparable to the standards of service on long continental connections in the early days of their development 15 years ago. The costs of intercontinental service are materially higher than the costs for the longest connections on continental networks even where distances are similar. These facts indicate that to a large extent the future development of intercontinental telephony is dependent upon a continuance of those brilliant advances in the communication art which have made our present intercontinental circuits possible. Further technical developments making it possible to improve the quality and reduce the cost of intercontinental services will, as they become available, have a tremendous effect.

The development of overseas services of large magnitude will require a closer coordination between the telephone plants and practices of the various continental wire networks than exists at the present time. The circuits which will form the connecting link between the subscriber and the terminals of the overseas circuits must have such transmission characteristics as to provide for satisfactory operation of the complete connection, including the overseas circuit and the continental extensions at both ends. In one respect this will impose requirements more severe than necessary where continental service alone is in question. We refer to the velocity of transmission which determines the elapsed time between the speaking of a word at one end of a circuit and its reception at the other end. Losses of power and electrical distortions in circuits may, within limits, be compensated for but time once lost in the propagation of the conversation over the circuit cannot be regained.

Equally important is the closer coordination of operating methods and commercial practices in so far as they affect intercontinental communications. It is natural that continental telephone networks developing more or less independently should represent somewhat different solutions of the operating and commercial problems involved in giving telephone service. Intercontinental service brings new problems and requires the development of new operating methods and new commercial practices designed to simplify and expedite the handling of these connections.

In the closer cooperation between telephone administrations and the consideration of their joint problem which comes with the development of intercontinental telephone service, it is increasingly important that they have commonly accepted methods of measuring and of expressing all of the quantities affecting the types and grade of service to be given. The International Advisory Committee on Long Distance Telephony (The C. C. I.) which has been active in facilitating the cooperation of European administrations in the improvement of international telephony in Europe, has included in its program the development of internationally accepted terms and units which will be of help to the nations of the world in their rendering of intercontinental service.

The commercial success of an intercontinental telephone service depends upon the existence at its terminals of wire networks by means of which large numbers of telephone subscribers can be given connection to the intercontinental circuits. The ultimate development of a really universal service depends, therefore, in part on the creation of large national or continental wire telephone networks in the areas where these do not now exist. It is only in so far as this takes place that it will become practical to realize world-wide telephony.

As world-wide telephony overcomes these obstacles and develops in magnitude and completeness, what form will the system take? It is, of course, too early to give a categorical answer to this question, but present development gives us some indications from which to judge the future.

Considering first the field of use of the various types of circuit, it seems that radio circuits will for a long time fill an important field in the provision of intercontinental circuits. This is particularly true of the short-wave systems which seem to be best adapted for pioneer work such as is going on at the present time on light traffic routes. Here their imperfect reliability is more than offset by their flexibility and relative economy. Long-wave radio will, no doubt, continue to be valuable for certain routes where the direction of transmission is east and west and at a high latitude, but the limit, which is apparently inherent, on the total number of circuits of this type that can be used simultaneously in the world would seem to prevent them from supplying any large part of the world's future needs for intercontinental circuits. Radio telephony, both short-wave and long-wave, must compete for wave-lengths with other forms of radio service. It is evident from the important part that radio must continue to play in world telephony that the increasing needs for wave-lengths for this rapidly growing service will require special consideration in future international radio conferences. Wire circuits which in the present pioneer stage are just beginning to enter the scene, will undoubtedly become more important for the principal circuit groups as intercontinental communication develops.

As the amount of intercontinental traffic builds up, and as the technical form and best routing of the telephonic relations between continents become established, it is to be expected that experience will show the advisability of adopting a fundamental routing and switching plan similar to those plans which have already been considered and put into use for some of the large continental networks. It is now too early to suggest in any detail the form of such a switching plan, but it seems that while political considerations may temporarily affect the form of the network, ultimately the requirements of economy and good service which have determined the form of continental plans now in use will be weighty factors in the planning of a world-wide fundamental switching plan. This gives a clue to some of its characteristics.

The splitting of circuit time between different terminal wire networks, while a valuable expedient for offering service under pioneer conditions, will naturally disappear generally as sufficient traffic develops to justify a full time circuit on a given route. This may be expected to be followed by the development of small circuit groups between the more important continental telephone networks, each group being operated as a unit between fixed terminal points. The advantages, while circuit groups are small, of concentrating intercontinental traffic as far as possible between the same terminals rather than diverting it to individual circuits between different terminal points, are great. As an illustration, between the United States and Europe where the volume of traffic has already led to the development of a circuit group, the operation of the six circuits now contemplated as a unit is estimated to give for the same grade of service a capacity one third greater than would be afforded by six separate circuits of the same character between different terminal points. The inclusion in one group of circuits of three different types, short-wave radio, long-wave radio and cable, will afford a continuity of service and an insurance against interruption far beyond what could be achieved with single circuits. A single circuit group between the two continental networks is also advantageous from a service standpoint because of the simpler operating arrangements.

While the first stage in the development of intercontinental business appears to indicate the concentration of intercontinental traffic in so far as the extent of continental networks makes this practicable, a second stage in the development will naturally be the establishment of additional circuit groups between other points of the networks. This becomes economical and desirable from a service standpoint when the original circuit group becomes large enough to permit of subdivision without great loss in efficiency and when the amount of traffic which can be conveniently handled through additional points in the continental wire networks is sufficient to fill the time of a group of several circuits. Hence, it is to be expected that the ultimate switching plan for world-wide telephony will include between large continental networks, such for example as exist in Europe and in the United States, a number of groups of overseas circuits between different terminal points selected so as to handle the traffic most conveniently and economically. Such a plan would also necessarily provide arrangements for the use of alternate routes. In any case the ultimate best plan from the standpoint of service and economy will depend upon the volume of traffic.

The technical achievements which have made possible the linking together of the continents of the earth with telephone circuits are in a high degree romantic. What may be accomplished for the benefit of mankind by the continued development of this world-wide telephone network depends upon what is said over the telephone circuits. It is the hope of scientists and engineers who have been engaged in this work that this closer knitting together of the nations and races of the world will be a great contribution to international friendship and good will. Mr. Coolidge, President of the United States in 1927, on the occasion of the opening of the telephone connection between the United States and Spain, gave apt expression to this thought when he said, "I believe it to be true that when two men can talk together the danger of any serious disagreement is measurably lessened, and that what is true of individuals is true of nations. The international telephone which carries the warmth and the friendliness of the human voice will always correct what might be misinterpreted in the written word."

APPENDIX I

Intercontinental Telephone Circuits of the World
January 1, 1932

Existing Circuits[1]

Circuit Group Index No.	Circuit Designation	Ownership	Distance Kilometers*	Service Date	Rates[2]	Time Sharing Arrangements
	North America—Europe					
101	(1) London—New York (Long Wave)	British P. O.—Am. Tel. and Tel. Co.	5550	1/ 7/27	$30.00	Not shared
	(2) London—New York	British P. O.—Am. Tel. and Tel. Co.	5550	6/ 6/28	$30.00	Not shared
	(3) London—New York	British P. O.—Am. Tel. and Tel. Co.	5550	6/ 1/29	$30.00	Not shared
	(4) London—New York	British P. O.—Am. Tel. and Tel. Co.	5550	12/ 1/29	$30.00	Not shared
	North America—South America					
201	Buenos Aires—New York	Compania Internacional de Radio (I. T. and T. Co.)—Am. Tel. and Tel. Co.	8580	4/ 3/30	$30.00	At New York with 202
202	New York—Rio de Janeiro	Am. Tel. and Tel. Co.—Companhia Radio Internacional do Brasil (I. T. and T. Co.)	7810	12/18/31	$30.00	At New York with 201 At Rio de Janeiro with 802
	Europe—South America					
301	Buenos Aires—London	Compania Internacional de Radio (I. T. and T. Co.)—British P. O.	11140	12/12/30	£ 6/0	At Buenos Aires with 307, 802 At London with 302, 303
302	Buenos Aires—London	Transradio Internacional, Compania Radiotelegrafica Argentina S. A.—British P. O.	11140	12/12/30	£ 6/0	At Buenos Aires with 305, 308, 312, 801 At London with 301, 303

---

* 1 kilometer ≈ .6 mile.

[1] See notes at the end of this appendix.

APPENDIX I (*Continued*)

Circuit Group Index No.	Circuit Designation	Ownership	Distance Kilometers	Service Date	Rates[2]	Time Sharing Arrangements
	EUROPE—SOUTH AMERICA (*cont.*)					
303	London—Rio de Janeiro	British P. O.—Companhia Radiotelegraphica Brasileira (Transradio)	9290	5/21/31	£ 6/0	At London with 301, 302 At Rio de Janeiro with 306, 309, 314, 801
304	Buenos Aires—Paris....	Compania Internacional de Radio (I. T. and T. Co.)—Cie. Gle. de T. S. F.	11060	6/11/30	150 G. F.	At Buenos Aires with 311 At Paris with 305, 306
305	Buenos Aires—Paris....	Transradio Internacional Compania Radiotelegrafica Argentina S. A.—Cie. Gle. de T. S. F.	11060	2/ 1/29	150 G. F.	At Buenos Aires with 302, 308, 312, 801 At Paris with 304, 306
306	Paris—Rio de Janeiro...	Cie. Gle. de T. S. F.—Companhia Radiotelegraphica Brasileira (Transradio)	9170	3/31/30	750 F.	At Paris with 304, 305 At Rio de Janeiro with 303, 309, 314, 801
307	Berlin—Buenos Aires...	German P. O.—Compania Internacional de Radio (I. T. and T. Co.)	11920	9/10/30	120.00 M.	At Berlin with 308, 309 At Buenos Aires with 301, 802
308	Berlin—Buenos Aires...	German P. O.—Transradio Internacional Compania Radiotelegrafica Argentina S. A.	11920	12/10/28	120.00 M.	At Berlin with 307, 309 At Buenos Aires with 302, 305, 312, 801
309	Berlin—Rio de Janeiro..	German P. O.—Companhia Radiotelegraphica Brasileira (Transradio)	10030	3/21/30	120.00 M.	At Berlin with 307, 308 At Rio de Janeiro with 303, 306, 314, 801
310	Berlin—Maracay, Ven. .	German P. O.—Venezuelan Govt.	8300	9/13/31	120.00 M.	Not shared
311	Buenos Aires—Madrid. .	Compania Internacional de Radio (I. T. and T. Co.)—Compagna Telephonica Nationale d'Espagna (I. T. and T. Co.)	10060	10/12/29	150 G. F.	At Buenos Aires with 304 At Madrid with 313
312	Buenos Aires—Madrid. .	Transradio Internacional Compania Radiotelegrafica Argentina S. A.—Compania Transradio Espanola (Transradio)	10060	1929	150 G. F.	At Buenos Aires with 302, 305, 308, 801 At Madrid, with 314

APPENDIX I (*Continued*)

Circuit Group Index No.	Circuit Designation	Ownership	Distance Kilometers	Service Date	Rates[2]	Time Sharing Arrangements
	EUROPE—SOUTH AMERICA (*cont.*)					
313	Madrid—Santiago	Compagna Telephonica Nationale d'Espagna (I. T. and T. Co.)—Compania Internacional de Radio, S. A. (I. T. and T. Co.)	10700	4/17/31	165 G. F.	At Madrid with 311 At Santiago with 803, 804
314	Madrid—Rio de Janeiro	Compania Transradio Espanola—Companhia Radiotelegraphica Brasileira (Transradio)	8140	1930		At Madrid, with 312 At Rio de Janeiro with 303, 306, 309, 801
	EUROPE—AFRICA					
401	Casablanca—Paris	Moroccan Tel. & Tel. Admin.—Cie. Gle. de T. S. F.	1820	11/ 3/30	111 F.	At Paris with 503
402	Madrid—Teneriffe	Compagna Telephonica Nationale d'Espagna (I. T. and T. Co.)—Compagna Telephonica Nationale d'Espagna (I. T. and T. Co.)	1820	1/22/31	30 Pesetas	At Madrid with Madrid-Mallorca
	EUROPE—ASIA—OCEANIA					
501	London—Sydney	British P. O.—Amalgamated Wireless Australasia	17000	4/30/30	£ 6/0	Note 3
502	(1) Amsterdam—Bandoeng	Neth. Govt.—Neth. Indies Tel. Admin.	11740	1/ 8/29	33 Florins	At Amsterdam—Not shared At Bandoeng—Note 5
	(2) Amsterdam—Bandoeng	Neth. Govt.—Neth. Indies Tel. Admin.	11740	12/ 1/29	33 Florins	At Amsterdam—Not shared At Bandoeng—Note 5
503	Paris—Saigon	Cie. Gle. de T. S. F.—Cie. Gle. de T. S. F.	10120	4/11/30	450 F.	At Paris with 401
504	Bandoeng—Berlin	Neth. Indies Tel. Admin.—German P. O.	10830	12/29/29	96.00 M.	At Bandoeng—Note 5 At Berlin with 505
505	Bangkok—Berlin	Siamese Govt.—German P. O.	8610	4/15/31	96.00 M.	At Bangkok—Note 3 At Berlin with 504

APPENDIX I (*Continued*)

Circuit Group Index No.	Circuit Designation	Ownership	Distance Kilometers	Service Date	Rates ?	Time Sharing Arrangements
	NORTH AMERICA—ASIA OCEANIA					
601	Honolulu—San Francisco	Mutual Tel. Co.—Am. Tel. and Tel. Co.	3850	12/23/31	$21.00	Not shared
	NORTH AMERICA					
701	Hamilton (Bermuda)—New York	Imp. & Int. Comm. Ltd.—Am. Tel. and Tel. Co.	1280	12/21/31	$15.00	Not shared
	SOUTH AMERICA					
801	Buenos Aires—Rio de Janeiro	Transradio Internacional Compania Radiotelegrafica Argentina S. A.—Companhia Radio telegraphica Brasileira (Transradio)	1970	1931		At Buenos Aires—with 302, 305, 308, 312 At Rio de Janeiro—with 303, 306, 309, 314
802	Buenos Aires—Rio de Janeiro	Compania Internacional de Radio (I. T. and T. Co.)—Companhia Radio Internacional do Brazil (I. T. and T. Co.)	1970	12/12/31	75 G. F.	At Buenos Aires with 307, 301 At Rio de Janeiro with 202
803	Bogota—Santiago	All America Cables Inc. (I. T. and T. Co.)—Compania Internacional de Radio S. A. (I. T. and T. Co.)	4260	8/ 1/31	105 G. F.	At Santiago with 313
	ASIA—OCEANIA					
901	Sydney—Wellington	Australian P. O.—New Zealand Govt.	2210	11/25/30		Note 3
902	Bandoeng—Bangkok	Neth. Indies Tel. Admin.—Siamese Govt.	2360	4/15/31	37.50 Florins	At Bandoeng—Note 5 At Bangkok—Note 3
903	Bandoeng—Sydney	Neth. Indies Tel. Admin.—Australian P. O.	5460	12/23/30	73.00 Florins	At Bandoeng—Note 5 At Sydney—Note 3
904	Bandoeng—Medan (Sumatra)	Neth. Indies Tel. Admin.—Neth. Ind. Tel. Admin.	1410	9/16/31	12.00 Florins	At Bandoeng—Note 5

APPENDIX I (*Continued*)

INTERCONTINENTAL TELEPHONE CIRCUITS OF THE WORLD
JANUARY 1, 1932

Proposed Circuits[1]

Circuit Group Index No.	Circuit Designation	Ownership	Distance Kilometers	Service Date	Rates[2]	Time Sharing Arrangements
	NORTH AMERICA—EUROPE					
101	(5) London—New York (Long Wave)........	British P. O.—Am. Tel. and Tel. Co.	5550	1934	$30.00	Not shared
	(6) London—New York (Cable).............	Note 3—Am. Tel. and Tel. Co.				Not shared
102	London—Montreal.....	British P. O.—Canadian Marconi Ltd.	5220	1932	£ 6/0	At London—with 404 At Montreal—Not shared
103	Berlin—Mexico City....	German P. O.—Mexican Govt.	9720			Note 3
	NORTH AMERICA—SOUTH AMERICA					
203	Lima—New York......	Compania Peruana de Telefonos, Limitada (I. T. and T. Co.)—Am. Tel. and Tel. Co.	5920	1932		Note 3
204	Bogota—Miami........	Compania Telefonica Central (Assoc. T. & T. Co.)—Am. Tel. and Tel. Co.	2440	1932		Note 3
205	Maracay—Miami......	Compania Annonima Nacional Telefonos de Venezuela—Am. Tel. and Tel. Co.	2190	1932		Note 3

APPENDIX I (*Continued*)

Circuit Group Index No.	Circuit Designation	Ownership	Distance Kilometers	Service Date[4]	Rates[2]	Time Sharing Arrangements
	EUROPE—SOUTH AMERICA					
315	Madrid—Rio de Janeiro	Compagna Telephonica Nationale d'Espagna (I. T. and T. Co.)—Companhia Radio Internacional do Brasil (I. T. and T. Co.)	8140	1932	150 G. F.	At Madrid with 311, 313 At Rio de Janeiro with 202, 802
316	Brussels—Buenos Aires	Belgian Govt.—Transradio Internacional Compania Radiotelegrafica Argentina S. A.	11320	1932	150 G. F.	At Brussels—with 406 At Buenos Aires with 302, 305, 308, 312, 801
	EUROPE—AFRICA					
401	Casablanca—Paris	Moroccan Tel. & Tel. Admin.—French P. T. T.	1820			Not shared
403	Capetown—London	Overseas Com. Co. of So. Africa—British P. O.	9680	1932		At London—with 301, 302, 303
403A	Johannesburg—London	South African P. O.—British P. O.	9080			Note 3
404	Cairo—London	Marconi Radio Tel. of Egypt—British P. O.	3510	1932		At London—with 102 At Cairo—Note 3
405	(1) Algiers—Paris	French P. T. T.—French P. T. T.	1350	1932		Not shared
	(2) Algiers—Paris	French P. T. T.—French P. T. T.	1350	1932		Not shared
406	Brussels—Leopoldville	Belgian Govt.—Note 3	6000	1932		At Brussels—with 316
407	Berlin—Cairo	German P. O.—Marconi Radio Tel. of Egypt	2910			Note 3
408	Paris—Tananarive	French P. T. T.—French P. T. T.	8760			Note 3
409	Cairo—Paris	Marconi Radio Tel. of Egypt—French P. T. T.	3210			Note 3

APPENDIX I (*Continued*)

Circuit Group Index No.	Circuit Designation	Ownership	Distance Kilometers	Service Date[4]	Rates[2]	Time Sharing Arrangements
	EUROPE—ASIA—OCEANIA					
502	(3) Amsterdam—Bandoeng	Neth. Govt.—Neth. Indies Tel. Admin.	11740		33 Florins	Note 3
506	Bombay—London	Indian Radio Teleg. Co.—British P. O.	7300			At London—with 101
507	London—Singapore	British P. O.—Imp. & Int. Comm. Ltd.	12390			Note 3
508	London—Tokyo	British P. O.—Japanese Govt.	9560			Note 3
509	Hong Kong—London	Hong Kong Tel. Co.—British P. O	9630			Note 3
	NORTH AMERICA—ASIA—OCEANIA					
602	Manila—San Francisco	Philippine L. D. Tel. Co. (Assoc. T. & T. Co.)—Am. Tel. and Tel. Co.	11220			Note 3
603	San Francisco—Tokyo	Am. Tel. and Tel. Co.—Japanese Govt.	8260			Note 3
604	San Francisco—Sydney	Am. Tel. and Tel. Co.—Australian P. O.	11950			Note 3
	NORTH AMERICA					
702	Miami—Tegucigalpa (Honduras)	Am. Tel. and Tel. Co.—Tropical Radio Teleg. Co.	1480	1932		At Miami with 703, 704, 705
703	Miami—Managua (Nicaragua)	Am. Tel. and Tel. Co.—Tropical Radio Teleg. Co.	1610	1932		At Miami with 702, 704, 705

APPENDIX I (*Continued*)

Circuit Group Index No.	Circuit Designation	Ownership	Distance Kilometers	Service Date[4]	Rates[2]	Time Sharing Arrangements
	NORTH AMERICA (*cont.*)					
704	Miami—San Jose (Costa Rica)	Am. Tel. and Tel. Co.—Tropical Radio Teleg. Co.	1800	1932		At Miami with 702, 703, 705
705	Miami—Panama	Am. Tel. and Tel. Co.—Tropical Radio Teleg. Co.	1870	1932		At Miami with 702, 703, 704
706	Juneau—San Francisco	Note 3—Am. Tel. and Tel. Co.	2530			
	SOUTH AMERICA					
804	Lima—Santiago	Compania Peruana de Telefonos, Limitada (I. T. & T. Co.)—Compania Internacional de Radio, S. A. (I. T. & T. Co.)	2470			At Lima—Note 3 At Santiago with 803, 313
	ASIA—OCEANIA					
905	Bangkok—Calcutta	Siamese Govt.—Indian Radio Co.	1620			Note 3

Notes

1. The table includes only circuits over 1000 km. in length. All circuits are short wave radio circuits except as noted in the New York—London group.
2. Rates given are for an initial 3 minute period. Abbreviations used are as follows: G. F. = Gold Francs, F. = French Francs, M. = Marks.
3. Definite information is not available.
4. Probable service date.
5. There are five transmitters at Bandoeng which serve the six circuits shown in the table on a shared basis.

APPENDIX II

Selected Bibliography

The following bibliography does not purport to be complete but is selected from among the articles bearing on the subject of "World-Wide Telephony" which have come to the attention of the authors.

Angwin, A. S. Chairman's Address, Wireless Section, *Journal of Institution of Electrical Engineers*, December 1931, p. 17–35.

Bailey, Austin, Dean, S. W. and Wintringham, W. T. Receiving System for Long-wave Transatlantic Radio Telephony. Institute of Radio Engineers, *Proceedings*, Vol. 16, No. 12, December 1928, p. 1645–1705.

Beer, C. A. and Evans, G. T. Transatlantic Telephony; the Post Office Differential Voice-Operated Anti-Singing Equipment. *Post Office Electrical Engineers Journal*, Vol. 20, April 1927, p. 65–72.

Blackwell, O. B. Transatlantic Telephony—the Technical Problem. American Institute of Electrical Engineers, *Journal*, Vol. 47, No. 5, May 1928, p. 369–373. *Bell System Technical Journal*, Vol. 7, April 1928, p. 161–167.

Blackwell, O. B. The Time Factor in Telephone Transmission. *Bell System Technical Journal*, Vol. 11, No. 1, January 1932, p. 53–66. American Institute of Electrical Engineers, *Transactions*, Vol. 51, No. 1, March 1932.

Buttner, H. H. Role of Radio in the Growth of International Communication. *Electrical Communication*, Vol. 9, No. 4, April 1931, p. 249–254. Institute of Radio Engineers, *Proceedings*, Vol. 19, No. 1, January 1931, p. 51–58.

Craemer, P. Weltfernsprechnetz in seinen geographischen Bedingtheiten. World Engineering Congress. *Proceedings*, Tokio, 1929, Vol. 19, p. 497–505.

Craemer, P. Weltfernsprechverkehr. *Elektrotechnische Zeitschrift*, Vol. 50, No. 27, July 4, 1929, p. 959–963.

Craemer, P., and Franke, A. Länderkarten des Europäischen Fernsprechnetzes. Sonderheft des "Europäischer Fernsprechdienst." May 1931.

Erikson, P. E. Service Telephonique International. *Journal Telegraphique*, Vol. 55, No. 12, December 1931, p. 381–385.

Espenschied, Lloyd, Anderson, C. N., and Bailey, Austin. Transatlantic Radio Telephone Transmission. Institute of Radio Engineers, *Proceedings*, Vol. 14, No. 1, February 1926, p. 7–56. *Bell System Technical Journal*, Vol. 4, No. 3, July 1925, p. 459–507.

Espenschied, Lloyd and Wilson, W. Overseas Radio Extensions to Wire Telephone Networks. Institute of Radio Engineers, *Proceedings*, Vol. 19, No. 2, February 1931, p. 282–302. *Bell System Technical Journal*, Vol. 10, No. 2, April 1931, p. 243–264.

Gherardi, Bancroft and Jewett, F. B. Telephone Communication System of the United States. World Engineering Congress Proceedings, Tokio, 1929, Vol. 19, p. 413–496. *Bell System Technical Journal*, Vol. 9, No. 1, Jan. 1930, p. 1–100.

Heising, R. A. Production of Single Sideband for Transatlantic Radio Telephony. Institute of Radio Engineers, *Proceedings*, Vol. 13, No. 3, June 1925, p. 291–312.

Lee, A. G. Radio Communication Services of the British Post Office. Institute of Radio Engineers, *Proceedings*, Vol. 18, No. 10, October 1930, p. 1690–1731.

Lee, A. G. Transatlantic Telephony. *Telegraph and Telephone Journal*, Vol. 12, April 1926, p. 150–151; vol. 13, February 1927, p. 92–93. *Post Office Electrical Engineers Journal*, Vol. 19, April 1926, p. 74–75.

Litzinger, L. Möglichkeiten eines transkontinentalen Fernsprechweitverkehrs zwischen Europa und Asien. Europäischer Fernsprechdienst, No. 5, July 1927, p. 3–14.

Müller, E. Auf dem Wege zum Oceanfernsprechkabel. Europäischer Fernsprechdienst, No. 20, November 1930, p. 379–385.

Oswald, A. A. and Schelleng, J. C. Power Amplifiers in Transatlantic Radio Telephony. Institute of Radio Engineers, *Proceedings*, Vol. 13, No. 3, June 1925, p. 313–361.

Pilliod, J. J. Intercontinental Radiotelephone Service from the United States. American Institute of Electrical Engineers, *Transactions*, Vol. 51, No. 1, March 1932.

SHAUGHNESSY, E. H. Rugby Radio Station. *Post Office Electrical Engineers Journal*, Vol. 19, January 1927, p. 373–382. *Electrical Review*, Vol. 98, April 30, May 7, 1926, p. 716–719, 753–756. *Electrician*, Vol. 96, April 23, 1926, p. 468–469.

TOWNSHEND, H. International Telephony. *Telephone and Telegraph Journal*, Vol. 15, February 1929, p. 87–89; March 1929, p. 106–108; April 1929, p. 128–129; May 1929, p. 148–150.

Transoceanic Telephone Service; a symposium. Contents: Miller, T. G. General Aspects. Bown, Ralph. Short-Wave Transmission.* Oswald, A. A. Short-Wave Equipment.* Cowan, F. A. Short-Wave Stations. American Institute of Electrical Engineers, *Transactions*, Vol. 49, No. 2, April 1930, p. 621–643. **Bell System Technical Journal*, Vol. 9, April 1930.

WAGNER, K. W. Plan einer Fernsprechkabelverbindung zwischen Europa und Amerika. Reprint from Sitzungsberichten der Preussischen Akademie der Wissenschaften-Physikalisch-Mathematischen Klasse, Vol. 7, 1929.

WAGNER, K. W. Transozeanische Fernsprechkabel, *Elektrische Nachrichten Technik*, Vol. 6, No. 4, April 1929, p. 125–138.

WATERSON, K. W. Transatlantic Telephone Service—Service and Operating Features. American Institute of Electrical Engineers, *Journal*, Vol. 47, April 1928, p. 270–273. *Bell System Technical Journal*, Vol. 7, No. 2, April 1928, p. 187–194.

WILSON, W. AND ESPENSCHIED, LLOYD. Radio Telephone Service to Ships at Sea. *Bell System Technical Journal*, Vol. 9, No. 3, July 1930, p. 407–428.

WRIGHT, S. B. AND SILENT, H. C. The New York-London Telephone Circuit. *Bell System Technical Journal*, Vol. 6, No. 4, October 1927, p. 736–749.

Long Distance Telephone Circuits in Cable *

By A. B. CLARK and H. S. OSBORNE

This paper first very briefly reviews the history of long distance telephone cables in the United States. A statement is then given of the basis of the electrical design of present day cables, followed by a discussion of the standards applied to cable circuits and the application of cables to the telephone needs of the country. While the present system is satisfactory for the circuits now used in cable up to distances of 1800 miles (3000 kilometers) or more, it would not be satisfactory for the much greater distances expected for the future, both for continental and intercontinental service. The paper closes with a brief account of the progress which has been made in the development of a cable carrier telephone system which is expected to be satisfactory for any distances which may in the future be spanned by telephone circuits in cable.

The authors wish to acknowledge their appreciation of the assistance given them by a number of their associates in the American Telephone and Telegraph Company, particularly Messrs. L. G. Abraham and R. M. Goetchius.

DURING the last two decades there has been, in all countries which have a large telephone development, a remarkable increase in the use of long distance telephone circuits in cable resulting in the building up of such large networks of long distance telephone cables as those which today cover the continent of Europe and a part of the continent of North America. This paper discusses the technical problems encountered in this development and the solutions applied in the development of the long distance cable network of the United States of America, using this as an illustration because it is the telephone plant with which the authors are most familiar.

Important Cable Development Milestones in the United States

The early long distance telephone circuits in the United States were practically all open wire. In fact, great care was exercised in laying out open-wire circuits to eliminate the necessity for using even short stretches of cable. Cable began to be considered seriously for long distance service when loading became available. In 1902 the first commercially loaded cable circuits in the United States were installed between New York City and Newark, N. J., a distance of about 11 miles (17 kilometers). Other cables rapidly followed, until in 1906 loaded cables were installed between New York and New

---

* Presented at The International Electrical Congress, Paris, France, July 5–12, 1932.

Haven, a distance of 79 miles (127 kilometers), and between New York and Philadelphia, a distance of 87 miles (140 kilometers). These 1906 cables contained No. 14 A. W. G. conductors (1.6 millimeters in diameter) and were loaded with 250-millihenry inductance coils spaced about 6000 feet (1830 meters) apart. By means of these cables it was possible to obtain transmission equivalents low enough so that these circuits furnished what was then considered very good long distance service. These early cables consisted entirely of physical circuits.

In order to obtain more circuits from the same number of wires, much development effort was then spent on methods for phantoming cable circuits and loading the phantoms. A number of problems were encountered and solved, the most serious being the problem of avoiding undue crosstalk between the circuits due to fortuitous unbalances introduced at the loading points and between the cable conductors. The problem was finally solved and in 1910 a short cable was installed between Boston and Neponset, Mass., providing loaded phantom circuits as well as side circuits, thus increasing the number of circuits 50 per cent. This was followed by a rapid extension of the use of this type of cable.

The ultimate in large gauge loaded cables was achieved in 1914 when the installation was completed of underground cable from Boston to Washington, a distance of 450 miles (724 kilometers), New York City being about at the midpoint. Some of the conductors in the cables were No. 10 A. W. G. (2.6 millimeters in diameter) while others were No. 13 A. W. G. (1.8 millimeters in diameter). Most of the loading consisted of 200-millihenry coils on the sides and 135-millihenry coils on the phantoms, spaced 7400 feet (2255 meters) apart. The 1000-cycle losses per mile (per kilometer) on these circuits were .050 (.031) db for the sides and .040 (.025) db for the phantoms for No. 10 A. W. G. and .085 (.053) db and .070 (.043) db for the side and phantom, respectively, of No. 13 A. W. G.

The vacuum tube telephone repeater was demonstrated as a great success when the New York-San Francisco telephone line was officially opened January 25, 1915. When this device was applied to the then available loaded cable circuits various imperfections, unimportant on non-repeatered circuits, produced serious effects, some of which had already been encountered in the work leading up to the loading of the open wire transcontinental line. Among these were the impedance characteristics of the cable circuits which were irregular due in part to insufficient stability and uniformity in the capacitances of the individual loading sections. These impedance irregularities prevented good repeater balances being obtained and consequently restricted the amplification which could be utilized on two-wire circuits.

It was soon realized that even if the loading were made very uniform two-wire repeatered circuits would be restricted in their transmission ranges, partly because of the unbalances encountered at repeaters and partly because of the tendency of the circuits to crosstalk into each other. Experiments were therefore begun utilizing the four-wire circuit method. When four-wire circuits were first set up using the large gauge loaded circuits between Boston and Washington great difficulty was experienced in obtaining even reasonably uniform attenuation at different frequencies. It then became apparent that smoother (more uniform) circuits would be necessary. It also became evident that higher cutoff frequency and higher velocity loading was necessary, in order to widen the effective transmission band and reduce delay distortion and echo effects. As a first step in this direction a system of loading in which inductance coils of 175 millihenries on the side circuits, spaced 6000 feet (1830 meters) apart, was introduced, primarily for two-wire circuits. This was commonly referred to as medium-heavy loading, having a cutoff frequency of approximately 2800 cycles. Experiments on long four-wire circuits with this loading which were specially set up for test confirmed the previous ideas as to the seriousness of echo effects and delay distorton effects and made it apparent that a much lighter weight and higher cutoff loading would be necessary for great distances. Accordingly, a system of loading known as H-44-25 (6000 feet—1830 meters—spacing with 44-millihenry coils on the sides and 25-millihenry coils on the phantoms) was introduced. This was the beginning of modern long distance cable circuits in America.

Electrical Design of Toll Cable Circuits

It is impracticable in a short paper such as this to deal fully with all of the considerations which determine the electrical design of cable circuits. There are set down here, however, the important characteristics of the types of circuit which have been adopted for use in the United States and the reasons why certain arrangements were selected. Many specific designs which will meet the transmission objectives are, of course, possible. The designs selected have been based upon the aim of obtaining the desired results in the most economical manner, including the advantage which comes from concentrating on a small number of types of circuits, rather than on a larger number designed to meet accurately the requirements of different types of situations. In other countries where the ratios of costs for different parts of cable systems are different it is to be expected that the most economical designs will differ.

Cable Constants

The toll cable which is standard in the Bell System plant has capacitance of .062 microfarad per mile (.038 mf. per km.) for the side circuits and about .100 microfarad per mile (.062 mf. per km.) for the phantoms. There appears to be little to gain for voice-frequency circuits by varying materially from these capacitance figures.

With respect to size of wire, No. 19 A. W. G. conductors (.9-millimeter diameter) are well suited for both two-wire and four-wire circuit operation. No. 16 A. W. G. conductors (1.3-millimeter diameter) have been employed to a considerable extent in the past. In new cables conductors of this gauge are, in general, used only for relatively short non-repeatered circuits, or for program circuits. At the present time the possible economic advantage of using smaller sizes of wire than No. 19 A. W. G. is so small that it is considered to be outweighed, in general, by the greater complexity and variability of the circuits which would result.

Side Circuits and Phantoms

With the exception of program transmission circuits, multiple twin quads, utilizing both the side and phantom circuits, are used exclusively.

The ratio between the capacitances of phantoms and side circuits is about 1.6, while the ratio of resistances is 0.5. Because of this the phantoms which are loaded for the same cutoff frequency as the side circuits have lower attenuation and lower impedance than the side circuits. For repeatered circuits the phantoms and sides are so operated that they give substantially equal transmission results, this being desirable for flexibility reasons.

Two-Wire and Four-Wire Circuits

As is well known, two-wire repeatered circuits are more economical for the shorter distances while four-wire circuits are necessary to economically provide satisfactory transmission for longer distances. It is possible for terminating business to design two-wire cable circuits which will deliver good telephone service for distances of at least 900 miles (about 1500 kilometers). However, to meet the transmission standards current in the United States four-wire circuits are generally more practicable for distances more than two or three hundred miles (a few hundred kilometers).

Inductance and Spacing of Loading Coils

Theoretically the inductance and spacing of loading coils might be varied for each circuit length in order to obtain the most economical design. Studies which have been made, however, indicate that, taking into account the advantage of flexibility, it is desirable to use only two types of spacings of loading coils and only two general types of loading units. In the Bell System the toll loading coil spacings are 3000 feet (915 meters) and 6000 feet (1830 meters), loading with these spacings being designated B and H, respectively. For two-wire circuits, 88-millihenry loading coils for the sides and 50-millihenry coils for the phantoms are used with both spacings, giving loadings designated as B-88-50 and H-88-50. The choice between these is dictated by the repeater spacing. If less than about 45 miles (72 kilometers), H-88-50 loading is used; if greater, B-88-50 is used. Two-wire H-172-63 loading was used in the past for two-wire circuits but this has now been given up for new work in favor of the wider frequency band B and H-88-50 systems.

For four-wire circuits, as stated above, the standard loading is H-44-25 meaning, of course, 6000-foot (1830-meter) spacing of coils with 44-millihenry coils on the sides and 25-millihenry coils on the phantoms.

Important Transmission Characteristics of Loaded Cable Conductors

The characteristics of loaded cable circuits depend principally upon the electrical constants of the cable conductors, the inductance of the loading coils and their spacing. Some of the more important transmission characteristics of the loaded cable systems employing cables and loading coils of the type just described are given in Table I.

Spacing of Repeaters and Automatic Transmission Regulators

Repeaters are spaced as close to 50 miles (80 kilometers) apart as practicable. In the past variations upward from this to about 60 miles (100 kilometers) were allowed but it is now believed best to avoid such long spacings. Where the location of cities or other geographical situations make it desirable, spacings less than 50 miles (80 kilometers) are used.

Automatic transmission regulators are provided for circuits in aerial cables in excess of 50 to 100 miles (80 to 160 kilometers) in length and are preferably placed at every second repeater station. The devices are such, however, that satisfactory results may be obtained if regulators are as far apart as about 150 miles (250 kilometers), while under certain conditions circuit flexibility considerations call for regulators at adjacent repeater stations. With cable entirely underground transmission regulators will function satisfactorily from the electrical standpoint spaced as far apart as 300 miles (500 kilometers) although circuit flexibility considerations usually call for breaking circuits up into shorter regulator sections. In general, regulators are provided for all circuits in underground cable in excess of approximately 180 miles (300 kilometers) in length.

TABLE I

Characteristics of 19-Gauge Repeatered Toll Cable Circuits

	Two Wire				Four Wire 11-44-25	
	B-88-50		H-88-50			
	Side	Phantom	Side	Phantom	Side	Phantom
Characteristic Impedance—Ohms at 1000 Cycles . . .	1,560	930	1,120	670	800	450
Attenuation at 1000 Cycles						
db Per Kilometer at 55° F.	.17	.15	.22	.19	.30	.25
db Per Mile at 55° F.	.28	.23	.35	.30	.47	.39
Nominal Velocity—						
Kilometers Per Second. . . .	16,000	17,000	23,000	24,000	31,500	33,000
Miles Per Second	10,000	10,500	14,300	15,000	19,000	20,000
Cutoff Frequency of Loading—Cycles Per Second. . . .	5,600	5,900	4,000	4,200	5,600	5,900
Attenuation Change at 1000 Cycles—						
db Per Kilometer						
Aerial (55°–109° F.)	.019	.016	.025	.022	.034	.029
Underground (55°–73° F.)	.007	.006	.009	.007	.011	.010
db Per Mile						
Aerial (55°–109° F.)	.031	.026	.041	.035	.055	.046
Underground (55°–73° F.)	.011	.009	.014	.012	.018	.015

Gains of Repeaters

In four-wire circuits the spacing of the repeaters and their gains depends largely on "one-way circuit" considerations. The maximum levels are fixed by the repeater and loading coil capacities to handle speech waves without distortion, the lower levels being fixed primarily by noise. The lower levels also depend somewhat on crosstalk considerations, particularly crosstalk between circuits transmitting in opposite directions but this crosstalk is usually not controlling. The upper level limit used is + 10 db while the lower is − 24 db, these being referred to the level at the transmitting end of the toll circuit as zero.

In two-wire circuits, the above "one-way circuit" considerations are unimportant in determining the levels, these being largely fixed by "two-way" considerations. The important "two-way" considerations are crosstalk on the one hand and proper control of circulating currents which might cause singing or serious echoes on the other hand. These considerations keep the repeater gains so low that the limit of repeater and loading coil capacity is not important nor is the lower limit at which noise would become serious.

The most important crosstalk consideration on two-wire circuits is "near-end" crosstalk. The arrangement of repeater gains which gives lowest near-end crosstalk is one in which the output levels of the repeaters are the same throughout the circuit until the receiving end is reached, at which point the repeater gain is reduced or loss inserted as necessary to give the desired overall net loss for the circuit.

This arrangement is not best from the standpoint of the circulating currents, however. From this standpoint the best setup is one in which the repeater output levels "taper" considerably from the sending end to the receiving end of the circuit.

In the Bell System plant a compromise is made between these two considerations which calls for layout of gains about as follows: At the transmitting end the transmitting repeater gain is made such that the outgoing level is + 3 db. As the transmission passes through other repeaters the upper level is allowed to drop about ½ db per repeater for average temperature conditions. Finally at the receiving end of the circuit gain or loss is introduced to give the required net loss for the circuit. Of course, the application of the above rules for laying out repeater gains results in giving individual repeaters different gains in the two directions.

Smoothness of Impedance

For two-wire circuits it is important that the cables have a "smooth" impedance-frequency characteristic. To attain this result, loading coil inductances are held within close manufacturing tolerances while the cable capacitance variations are also carefully controlled. In the field care is, of course, taken with the spacing of the loading coils. Following are some representative figures for side circuits of fractional deviation from the average values per loading section or per loading coil:

	H Spacing	B Spacing
Representative deviation of cable capacitance *	.013	.018
Representative deviation of loading coil spacing *	.005	.005
Representative deviation of loading coil inductance *	.007	.007
Total deviation	.016	.020

* Representative deviation is the square root of the mean square of the individual deviations.

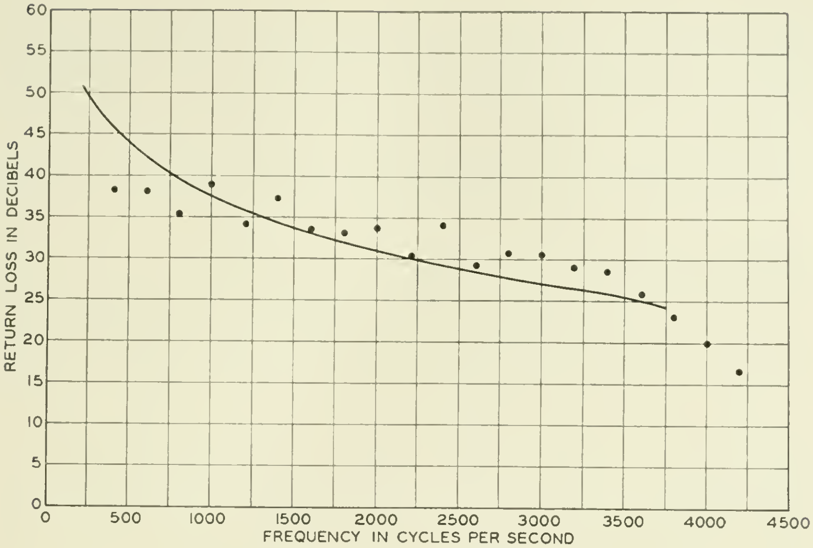


Fig. 1—Representative return losses on 19-ga. B-88-50 side circuits. 37 per cent of the measurements show lower return losses. ● 26 circuits Newark to Princeton.

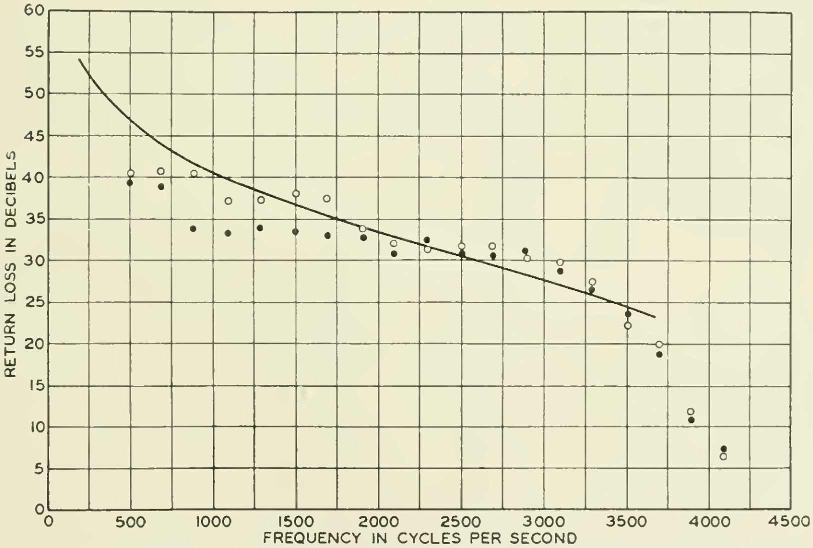


Fig. 2—Representative return losses on 19-ga. H-88-50 side circuits. 37 per cent of the measurements show lower return losses. o 26 circuits Princeton to Newark. ● 26 circuits Princeton to Philadelphia.

Representative return loss versus frequency curves are shown in Figs. 1 and 2 for B-88-50 and H-88-50 circuits, respectively, the points plotted being corresponding measurements on actual field facilities.

In the case of four-wire circuits, impedance irregularities are not so serious. However, for practical reasons, the same tolerances are usually followed for the several parts.

Control of Crosstalk

For two-wire circuits, the important crosstalk is near-end while for four-wire circuits it is far-end. For both of these, crosstalk between circuits within a single quad is greatest but crosstalk between circuits in different quads is also important. For two-wire circuits, in order to avoid long crosstalk exposures between any two circuits, it is the practice to carry three circuits together in a single quad only in a single repeater section, the circuits being systematically mixed at each repeater station. In the case of four-wire circuits this mixing is done only at the ends of regulator sections. In both outside cables and in the office cable, care is exercised to segregate the oppositely bound pairs of four-wire circuits because of the relatively large level differences.

In the outside cables control of crosstalk involves adjustments of the fortuitous unbalances in the loading units and unbalances between circuits in the loading sections. The following table shows the standards for phantom-to-side crosstalk expressed in decibels ordinarily worked to for the component parts of cable circuits:

	Two-Wire				Four-Wire	
	B-88		H-88		H-44-25	
	Avg.	Max.	Avg.	Max.	Avg.	Max.
Repeater section outside plant*	78		82		76	
Loading coil	96		96		96	
Cable proper per load section	93		94		94	
Toll office*	79		82		80	
Office wiring	80		86		82	
Office equipment	86		84		84	
Repeaters		91		91		74

* For two-wire circuits on the average these values must be decreased about 8 db to compare with overall value; for four-wire circuits about 9 db should be added to these values to compare with the overall.

Note: All values in table are for a frequency of 1250 cycles.

Control of Delay Distortion

In the case of two-wire circuits, since only relatively short distances are involved, delay distortion does not enter as a design factor. In the case of long four-wire circuits delay distortion is very important and to avoid this, very light weight and high cutoff loadings are used. Delay distortion may be reduced by employing correcting networks which introduce distortion counter to that introduced by the line. Such networks, in addition to their cost, have the disadvantage that they increase the total delay of the circuit. Their use is not necessary for the lengths of circuits and types of construction now used for message telephone circuits in the United States.

Performance Characteristics

Minimum Working Net Loss

In determining the numbers of circuits of different types of construction to be provided in a proposed cable for long distance work, it is necessary to take into account the limiting lengths for which the different types of circuit will meet the transmission requirements established for different conditions. This limitation for repeatered circuits is not set by attenuation but rather by singing margin, crosstalk or echoes. The lowest net loss at which a circuit equipped with repeaters may be operated without passing the limiting requirements for any of these characteristics is called the minimum net loss of the circuit. Since it is desirable to keep the net loss of a circuit at any time at least as great as the minimum net loss, an allowance for the probable circuit variations is added to the minimum net loss to obtain the minimum working net loss. This minimum working net loss is in general not directly proportional to the length of the circuit although in some cases within the important range of distances it can be considered to be proportional to a sufficient degree of accuracy.

It is evident that the minimum working net loss is a characteristic of fundamental importance in the design of new toll cables as well as in determining the operating limitations of existing cable circuits. Circuits having the same type of loading and the same spacing of repeaters may have widely different minimum net losses. This can come about because of differences in the accuracy with which crosstalk coupling is reduced in the manufacture and installation of the cable and its associated loading coils and apparatus, differences in the degree of uniformity of these characteristics and differences in the perfection of balance and matching of impedance between the cable conductors and the associated equipment. In the United States the study of the crosstalk, impedance and echo results obtained in various toll cables has led to the development of standard practices as to uniformity of construction and detailed specifications of the equipment with which the minimum net loss to be obtained with new toll cables can be closely predicted.

Performance Characteristics of Two-Wire Circuits

Fig. 3 shows, for B and H-88-50 two-wire circuits used exclusively for terminating business, the minimum working net losses for various distances as limited separately by crosstalk, echoes and singing. Of course, the lower the net loss the greater is the tendency toward excessive echoes, crosstalk or possibility of singing. For any given case the most exacting limitation controls the minimum working net loss.

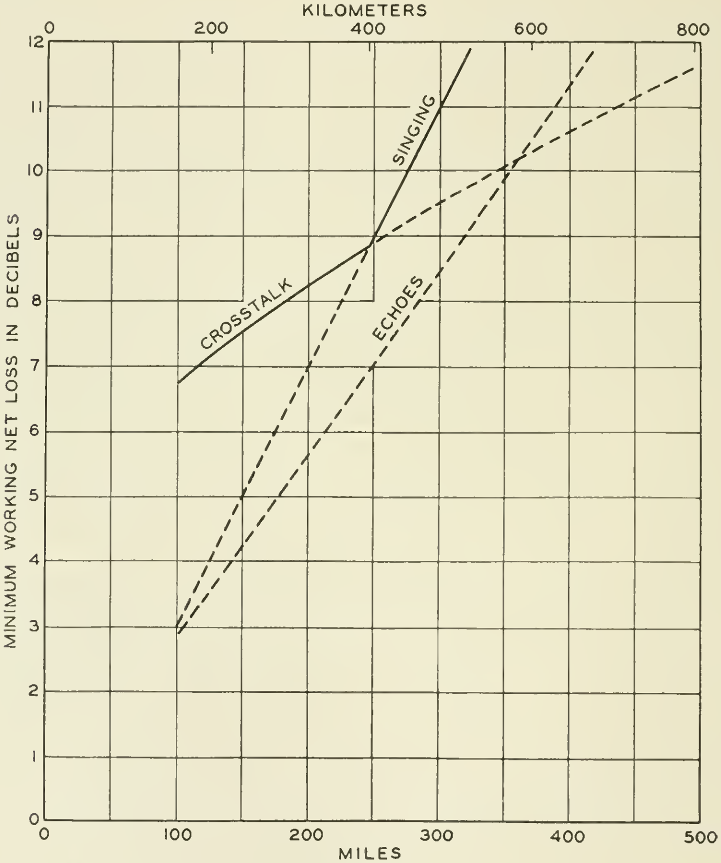


Fig. 3—Minimum working net loss of 19-ga. B and H-88-50 two-wire facilities versus circuit length for terminating business. These apply to either 50-mile (80 km.) sections of B-88-50 or 45-mile (72 km.) sections of H-88-50, whichever has the more severe limitations.

These curves include allowance for the unavoidable variations in the net loss which occur from time to time due to repeater battery variations, residual variations left over after pilot wire regulators have removed major transmission variations due to temperature changes in the cables, humidity effects in office wiring, etc.

Performance Characteristics of Four-Wire Circuits

Fig. 4 shows, for four-wire circuits used exclusively for terminating business, the minimum working net losses for various distances as limited separately by crosstalk and echoes. The possibility of singing does not enter as a limitation on these circuits. In these curves suitable allowance has also been made for the effect of the unavoidable transmission variations.

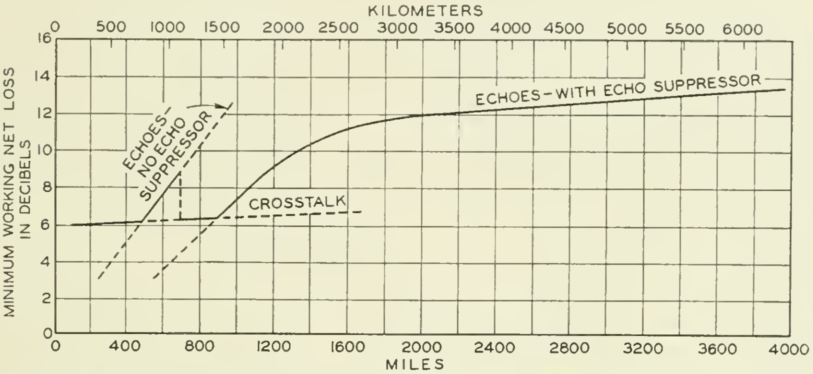


Fig. 4—Minimum working net loss of 19-ga. H-44-25 four-wire facilities versus circuit length for terminating business.

It will be observed that crosstalk is controlling only for the shorter distances up to a little over 400 miles (640 kilometers). Up to this length meeting proper crosstalk limits requires that the net loss be kept above about 6 db.

Echo constitutes the important limitation to the net loss of four-wire circuits. As is well known, echo suppressors go a long way toward eliminating echo effects but do not remove these effects, which remain the most important limiting factor on the longer circuits.

Transmission Requirements

The transmission requirements established for toll circuits are based on the provision of adequate transmission for the complete connection between any two points in the United States and the southern part of Canada, and also between any point of the country and the terminals of the intercontinental circuits. In order that this may be accomplished in an economical and orderly way a general toll switching plan has been adopted for the entire continental area. This plan establishes a fundamental basis for the routing of connections involving more than one toll circuit through the establishment of about 150 important switching centers to which all of the 2500 toll centers of the country will be directly connected. These 150 switching centers are interconnected by groups of high grade toll circuits either directly or for distant parts of the country through the intermediary of "regional centers" of which there are eight in the continental area.

By means of this plan it is possible to allocate each group of toll circuits to one of several broad classifications depending upon its position in the general toll switching plan, and to apply standard transmission requirements to each such broad classification. These requirements include the requirement that the effective net loss of all direct circuits shall not exceed 9 db, and that circuits designed for use in switched connections shall have minimum working net losses not exceeding 3 db for end links, 4 db for circuits between regional centers and 3.5 to 4 db for the remaining intermediate links interconnecting the important switching centers. When several circuits are connected together to form a long switched connection, the overall crosstalk effects are not appreciably increased over the effects of an individual circuit. Singing effects will usually not be limiting since long switched connections are seldom established without at least one intervening four-wire circuit. On the other hand echo effects increase fairly rapidly even with circuits equipped with echo suppressors and therefore, in selecting facilities which will meet the requirements of minimum working net loss previously specified, the echo effects are generally controlling on long connections. Since these circuits are also used for direct circuit connections, higher net losses, which will be satisfactory from the crosstalk and singing standpoints in the terminating condition, are obtained by adding pads at one or both terminals of the circuit for this condition.

A more complete statement of the transmission requirements applied to toll cable circuits in the United States is given in Table II. For convenience there are also given in this table the current transmission requirements for international circuits adopted by the C. C. I. In order to make the comparison as nearly as possible on a comparable basis the international circuit requirements of the C. C. I. are compared with the requirements for American toll circuits interconnecting two regional centers, and the requirements for the national terminal are compared with the American requirements for connections from the

TABLE II

Transmission Standards

Bell System		*C. C. I.*	
Number of Regional Centers —8		Number of European National Outlets —29	
NET LOSS			
Circuits Between Regional Centers (On effective transmission basis)		*International Circuits* (Based on 800-cycle transmission equivalent)	
Future Plant		*Future Plant*	
Terminating Business	9.0 db	Terminating or Switched Business	
Switched Business (Via Net Loss)*	3.0 db	Two-Wire	8.7 db
		Four-Wire	6.9 db
Existing Plant		*Existing Plant*	
	9.0–11.0	Two-Wire	11.3 db
	3.0–5 0 db	Four-Wire	9.6 db
MAXIMUM OVERALL NET LOSS BETWEEN SUBSCRIBERS†			
Regional Center Transmitting Loss	16 db	National Transmitting Loss	17 db
Regional Center Receiving Loss	12 db	National Receiving Loss	11 db
Circuit Between Regional Centers (Via Net Loss)	3–5 db	International Circuit	7–11 db
Total	31–33 db	Total	35–39 db
FREQUENCY BAND WIDTH			
250 to 2750 cycles Loss at extreme frequencies 10 db greater than at 1000 cycles. For narrower bands the effective transmission equivalent reflects the effect on transmission of band width.		300 to 2400 cycles Loss at extreme frequencies 8.7 db greater than at 800 cycles.	
DELAY DISTORTION			
Direct circuit connection (terminal circuit)—20 milliseconds. Via circuits (in switched connections) 10 milliseconds. (Differences between 1000 cycle delay and the highest frequency effectively transmitted.)		*Overall Connection* 30 milliseconds as the difference between the time of propagation for the highest frequency effectively transmitted and the time of propagation for 800 cycles.	
NON-LINEAR DISTORTION			
No fixed limits at present on message circuits but consideration is being given to methods of measuring and evaluating the impairment due to this effect.		No fixed limits at present on message circuits but consideration is being given to methods of measuring and evaluating the impairment due to this effect.	

* Via net loss is the loss which the circuit contributes to an overall connection when switched to toll lines at both ends.

† The regional center and national outlet transmitting and receiving losses are the maximum losses between the most distant subscriber in the particular area served by the regional center or national outlet and the particular regional center or national outlet.

TABLE II—(*Continued*)

Bell System	*C. C. I.*
TRANSMISSION VARIATIONS (AVERAGE OF THE TWO DIRECTIONS)	
250-mile (400-kilometer) circuits ± 1.5 db, 1000-mile (1600-kilometer) circuits ± 3.0 db (these limits are exceeded approximately 5 per cent of the time).	No limits.
ECHOES	
Echoes are limiting in accordance with curves representing the results of experience as to permissible loudness of echo without interference for different times of propagation. See page 129 of the "Red Book" of the C. C. I. On four-wire circuits echo suppressors are employed for all circuits used for switched business in excess of 270 miles (430 kilometers) of H-44-25 facilities. Four-wire circuits used for terminal business only are equipped with echo suppressors when they exceed approximately 650 miles (1000 kilometers) of H-44-25 facilities. Echoes on circuits so equipped are limiting as indicated in Fig. 4 of this paper.	Echoes are limiting in accordance with curves representing the results of experience as to permissible loudness of echo without interference for different times of propagation. See page 129 of the "Red Book" of the C. C. I.
SINGING MARGIN (TERMINAL CONDITION)	
10 db for two and four-wire circuits.	Two-wire 6.8 db. Four-wire 9.0 db.
CROSSTALK	
For quiet circuits: Average—70 db. Maximum—60 db. (Based on 1 per cent chance of exceeding 60 db.)	54 db at least.
MINIMUM WORKING NET LOSS*	
4 db.	No limits established.
TIME OF PROPAGATION	
For continental communication—250 milliseconds. Continental circuits in an intercontinental connection—100 milliseconds. For continental communication: Delay between echo suppressors—100 milliseconds.	For continental communication—250 milliseconds. Continental circuits in an intercontinental connection—100 milliseconds. For continental communication: Delay between echo suppressors—100 milliseconds.
NOISE (INCLUDING BABBLE) TERMINAL CONDITION MEASURED AT THE TOLL SWITCHBOARD	
+ 26 db (200 noise units) referred to reference noise (approximately 10 noise units). For noise values greater than above limit N. T. I.'s are applied.	*2 Millivolts (Approximately 160 Noise Units)* (Equivalent intensity of 800-cycle tone of the limiting voltage across a receiver whose impedance is adjusted to 600 ohms.)

* For more detailed discussion refer to text.

subscriber to a regional center. As there are eight regional centers in the United States and 29 national networks which constitute the European continental network, the areas involved in the comparison are not wholly comparable.

The application of these transmission requirements leads to the following approximate limits for the use of the different standard types of toll cable construction in the toll circuits of the United States:

19-gauge B and H-88-50 two-wire circuits	250 miles (400 kilometers) when used exclusively for terminating business
19-gauge B and H-88-50 two-wire circuits	135 miles (220 kilometers) for circuits used for switched business
19-gauge H-44-25 four-wire circuits without echo suppressors	650 miles (1050 kilometers) when used exclusively for terminating business
19-gauge H-44-25 four-wire circuits without echo suppressors	270 miles (430 kilometers) for circuits used for switched business

USE OF TOLL CABLES IN THE UNITED STATES

In the application of toll cables to meet the service requirements of the United States use is generally made of cables 2⅝ inches (67 millimeters) in outside diameter, although to some extent use has been made of cables 3⅛ inches (79 millimeters) in outside diameter. A large proportion of the new toll cables contain one of the combinations of gauges indicated in the following table:

TABLE III

	16-Gauge Quads	16-Gauge Pairs	19-Gauge Quads	22-Gauge Quads
Cable 1......	0	6	148	1
Cable 2......	19	6	114	1

The present toll cable routes in the United States are indicated on the map shown in Fig. 5, together with probable future extensions. The existing network includes 13,000 miles (21,000 kilometers) of route, 21,000 miles (34,000 kilometers) of cable, and when fully equipped, 5,000,000 miles (8,000,000 kilometers) of circuit.

The division of the cable between various types of construction is indicated in the following table:

Underground in ducts..................................	46%
Aerial..	50%
Buried underground....................................	4%

As an illustration of the practical application of toll cable, some information is given below regarding the makeup of some of the circuits out of New York City. There are associated with the New York toll board 235 groups of circuits. In order to avoid too cumber-

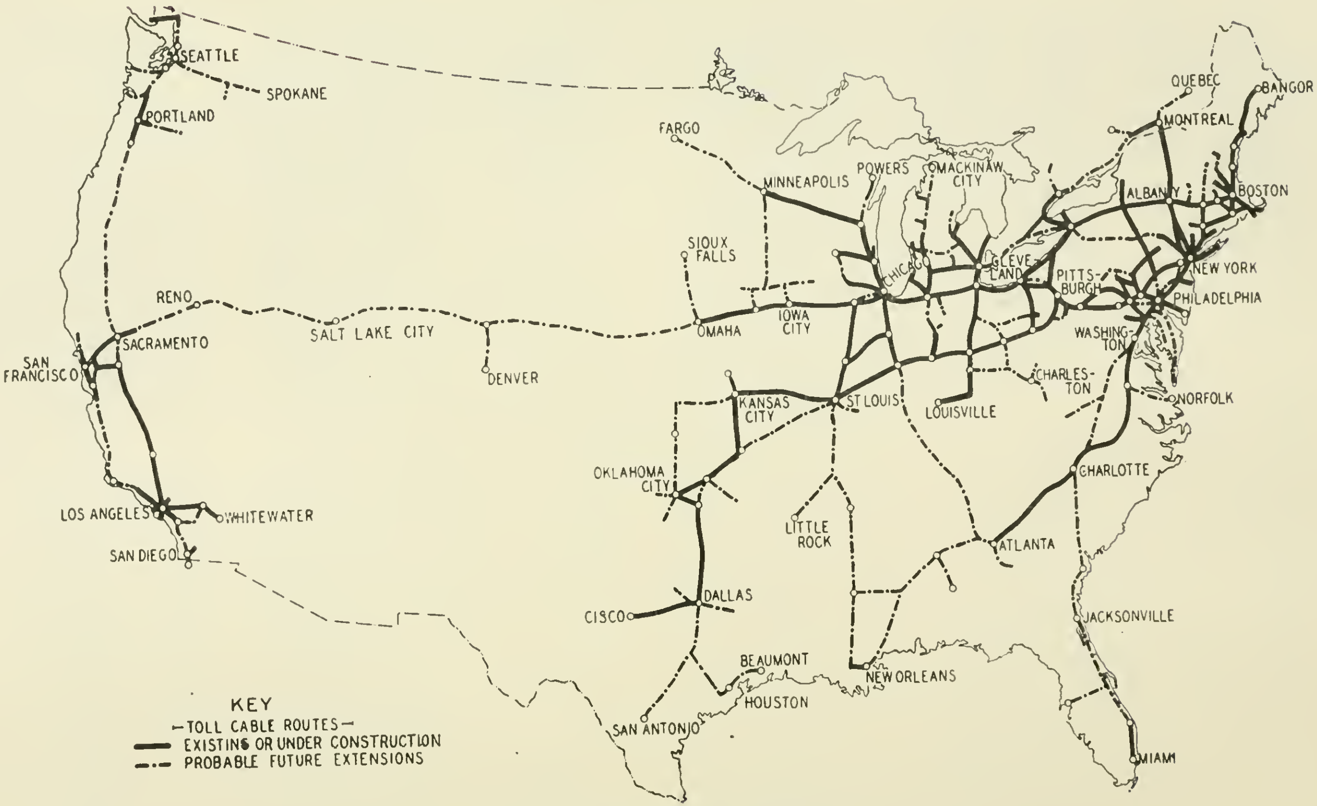


Fig. 5—Toll cable routes—Bell System.

some an illustration, we shall consider only circuits from New York to other regional centers and primary outlets. Chicago has a larger number of connections to these points but New York is chosen because of its interest in connection with international service. These circuits are indicated on the map in Fig. 6.

Of the 64 circuit groups to regional centers and primary outlets, 43 are less than 870 miles (1400 kilometers) in length, 95 per cent of the circuit mileage being cable. The makeup of a few of these groups chosen for the purpose of illustration is given in Table IV.

The remaining circuit groups indicated in Fig. 6, 870 miles (1400 kilometers) and more in length, and 21 in number, have the makeups and electrical characteristics indicated in Table V.

It will be noted that the average circuit in this classification is made up of about one half four-wire toll cable circuit and the other half of carrier telephone superimposed upon open wire. Since carrier circuits have, in general, electrical characteristics comparable to the four-wire cable circuits and have a relatively high velocity of propagation, the combination of four-wire with carrier results in very satisfactory electrical characteristics, even for the longest circuits.

A point of interest in connection with the table is the time of propagation. This is, in all cases, well within the provisional limit of 100 milliseconds adopted by the C. C. I. for the time of propagation of the continental terminating circuits of an intercontinental connection.

Future Requirements of Very Long Cable Circuits

At the present time cable up to lengths of 1800 miles (3000 kilometers) is used in the regular routine in the United States and gives a satisfactory performance. Tests have been made on longer four-wire cable circuits up to lengths exceeding 3600 miles (6000 kilometers). These tests show that for such lengths, and particularly for the much greater lengths which may result from the development of intercontinental telephone service, the present design of toll cable circuits would not be entirely satisfactory.

More Effective Echo Suppressors

Circuits 3600 miles (6000 kilometers) long when equipped with ordinary echo suppressors fail to give the net loss of 9 db which has been set up as a design objective. A more effective type of echo suppressor is necessary to work such a circuit at as low a net loss as 9 db. Experiments have been made with an echo suppressor of a type which changes its sensitivity automatically, depending upon the

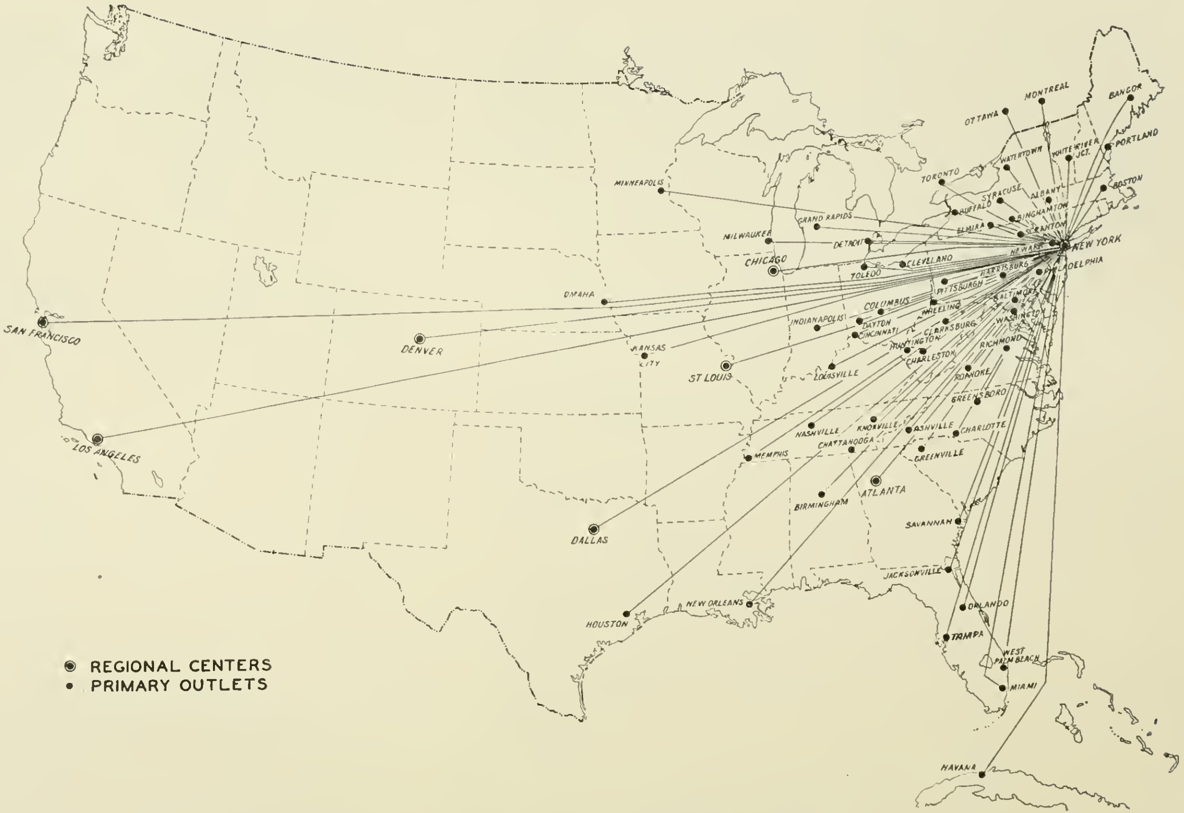


Fig. 6—Direct circuit groups from New York to regional centers and primary outlets.

TABLE IV

Typical Circuits from New York to Regional Centers and Primary Outlets Less than 870 Miles (1400 Kilometers)—January 1932

City	Class*	Number of Circuits	Length		Make-Up	Effective Net Loss—db	
			Kilometers	Miles		Terminal Condition	Via Condition
Philadelphia.....	Via	24	140	87	Four Wire H-44-25	6	0
	Term	299			Two Wire Older Types of Loading	12†	—
Boston..........	Via	17	367	228	Four Wire H-44-25	9	3
	Term	85			Two Wire Older Types of Loading	12†	—
Washington......	Via	7	359	223	Four Wire H-44-25	8	2
	Term	44			Four Wire H-44-25	8	—
Pittsburgh.......	Via	15	587	365	Four Wire H-44-25	8	2
	Term	24			Two Wire H-44-25	8	—
Baltimore.......	Via	6	299	186	Four Wire H-44-25	8	2
	Term	32			Four Wire H-44-25	8	—
Albany..........	Via	12	249	155	Two Wire H-44-25	8	2
	Term	23			Two Wire H-172-63	12†	—
Cleveland.......	Via	8	817	507	Four Wire H-44-25	9	3
	Term	18			Four Wire Older Type Loading	13†	—
Detroit.........	Via	9	1,090	677	Four Wire H-44-25	9	3
	Term	17			Four Wire Older Type Loading	12†	—
Toronto.........	Via	7	803	499	Four Wire H-44-25	9	3
	Term	9			Two Wire H-172-63	11†	—

* Via—Circuits used for switched business.
Term—Circuits used for terminal business only.

† With further growth of plant these circuits will be brought within present limits by reloading or by replacing with new circuits, limiting the older types of loading to shorter distances.

TABLE V

Direct Circuit Groups from New York to Regional Centers and Primary Outlets Excluding Those Less than 870 Miles (1400 Kilometers)—January 1932

City	Number of Circuits	Circuit Make-Up								Effective Net Loss db		1000-Cycle Delay Milliseconds
		Kilometers				Miles						
		Four Wire H-44-25	Type C Carrier	Other	Total	Four Wire H-44-25	Type C Carrier	Other	Total	Terminal Condition	Via Condition	
Atlanta	9	1,010	606	—	1,616	623	380	—	1,003	9	3	36
Birmingham	3	1,010	843	—	1,853	623	527	—	1,150	9	3	37
Chicago	59 *	1,400	—	—	1,400	—	—	—	—	9	3	46
Dallas	4	1,605	1,084	—	2,689	998	673	—	1,671	10	4	58
Denver	3	1,685	1,465	—	3,150	1,048	910	—	1,958	10	4	61
Havana	3	359	2,300	—	2,659	221	1,430	—	1,651	10	4	22
Houston	3	1,605	1,210	—	2,815	998	751	—	1,749	10	4	58
Jacksonville	8	1,010	—	632	1,642	623	—	398	1,021	9	3	37
Kansas City	5	2,010	—	—	2,010	1,250	—	—	1,250	10	4	67
Los Angeles	7	1,485	4,110	—	5,595	922	2,554	—	3,476	11	5	66
Memphis	3	1,100	863	—	1,960	684	534	—	1,218	9	3	40
Miami	19	1,010	1,215	—	2,225	622	759	—	1,382	9	3	38
Minneapolis	5	1,685	647	—	2,332	1,048	401	—	1,449	9	3	39
New Orleans	7	1,010	1,425	—	2,435	623	890	—	1,513	9	3	39
Omaha	2	2,175	—	—	2,175	1,350	—	—	1,350	10	4	72
Orlando	3	1,010	—	900	1,910	623	—	559	1,182	10	4	39
St. Louis	13	1,605	—	—	1,605	998	—	—	998	9	3	53
San Francisco	6	1,830	3,430	—	5,260	1,137	2,133	—	3,270	11	5	75
Tampa	5	1,010	931	—	1,941	623	585	—	1,208	9	5	37
Tulsa	2	1,605	703	—	2,308	998	438	—	1,436	10	4	56
West Palm Beach	9	1,010	1,105	—	2,115	623	690	—	1,313	9	3	38

* 45 Circuits for Terminal Business Only.

Note: All of these circuits are equipped with echo suppressors.

amount of noise on the circuit at any given time, the less noise the more sensitive the device. It has been found possible to adjust this device to sufficient sensitivity to permit working a 3600-mile (6000-kilometer) circuit with 9 db net loss. When noise is added to the circuit the sensitivity of the device, of course, diminishes but the added echo is largely masked by the increased noise.

Delay Distortion

For a 3600-mile (6000-kilometer) H-44-25 circuit the difference in the delay at 1000 cycles and at 3000 cycles amounts to about 0.025 second. Experiments which have been made on circuits introducing very little non-linear distortion indicate that this amount of delay distortion by itself is not particularly serious. However, on the long four-wire circuits where non-linear distortion is also present, the effect of delay distortion becomes more pronounced so that it becomes quite objectionable. Delay distortion correctors would therefore be required for H-44-25 circuits of this length, although for circuits of 1800 miles (3000 kilometers) they do not appear necessary.

Time of Propagation

When a long connection is built up using cable circuits, the delay proper, quite apart from delay distortion, becomes important. For a 3600-mile (6000-kilometer) length of H-44-25 circuit equipped with delay equalizers the time of propagation is about one-quarter of a second in each direction. This time of propagation is generally considered about all that should be tolerated and is the C. C. I. tentative limit for a complete connection.

Adverse Interaction of Two Echo Suppressors

When voice-operated devices are introduced on very long cable circuits another complication results. Assume, for example, that the 3600-mile (6000-kilometer) connection is made up of two links, each equipped with an echo suppressor of the usual type, either mechanical relay or vacuum tube operated. Assume that the echo suppressors are 1800 miles (3000 kilometers) apart, the delay between these devices being one-eighth of a second for each direction of transmission. When conversations are carried on over this circuit it is found that occasionally when the speakers at the two ends utter words at nearly the same instant both echo suppressors respond, each echo suppressor blocking one direction of the circuit. Consequently, certain words or parts of words are lost. In telephone conversations it is found that with such a circuit arrangement if the time of propagation in each direction between echo suppressors does not exceed about 0.1 second, the amount lost apparently is not a serious handicap. The C. C. I. has tentatively recommended that 0.1 second be taken as a limit towards which it is desirable to work if practicable.

Better Cable Circuits Desirable for Extreme Distances

These difficulties with the present type of cable construction all become more pronounced if, instead of a 3600-mile (6000-kilometer) circuit, consideration is given, for example, to possible future intercontinental circuits in all cable construction. Such a circuit between San Francisco and Istanbul, for example, would be about 10,000 miles (16,000 kilometers). The time of propagation for such a connection would be about .6 of a second in each direction so that two-way conversations would be seriously impeded. Serious difficulties would also be experienced with the voice-operated devices and because of the accumulated distortions, including non-linear effects. While, therefore, it is possible that a circuit of this sort could be used for two-way telephony between Istanbul and San Francisco, the imperfections of such a circuit would be so outstanding as to warrant a serious effort to develop something better if this could be done at a reasonable cost.

Telephone Carrier in Cable

In order to obtain better transmission results over very long cable circuits in an economical manner the development of a carrier system for cables has been actively undertaken. The development work of the Bell System has now been carried to the point where it seems assured that it will be successful and that telephone carrier will have an important field of use in long distance cables on heavy routes.

This carrier system uses non-loaded cable conductors whose velocity is very high as compared to voice-frequency loaded circuits, the effective circuit velocity including delays introduced by apparatus being about 100,000 miles (160,000 kilometers) per second. An experimental trial system has been set up by looping circuits back and forth in a cable so as to produce the equivalent of the system shown in Fig. 7. It will, of course, be understood that this figure represents merely the experimental setup and should therefore not be considered the ultimate in such matters as carrier channels per pair, repeater spacing, etc.

Talking tests which were made using this experimental setup showed very satisfactory quality of transmission and no appreciable interference between circuits In addition to testing each of the nine telephone circuits shown in the sketch, these nine circuits were con-

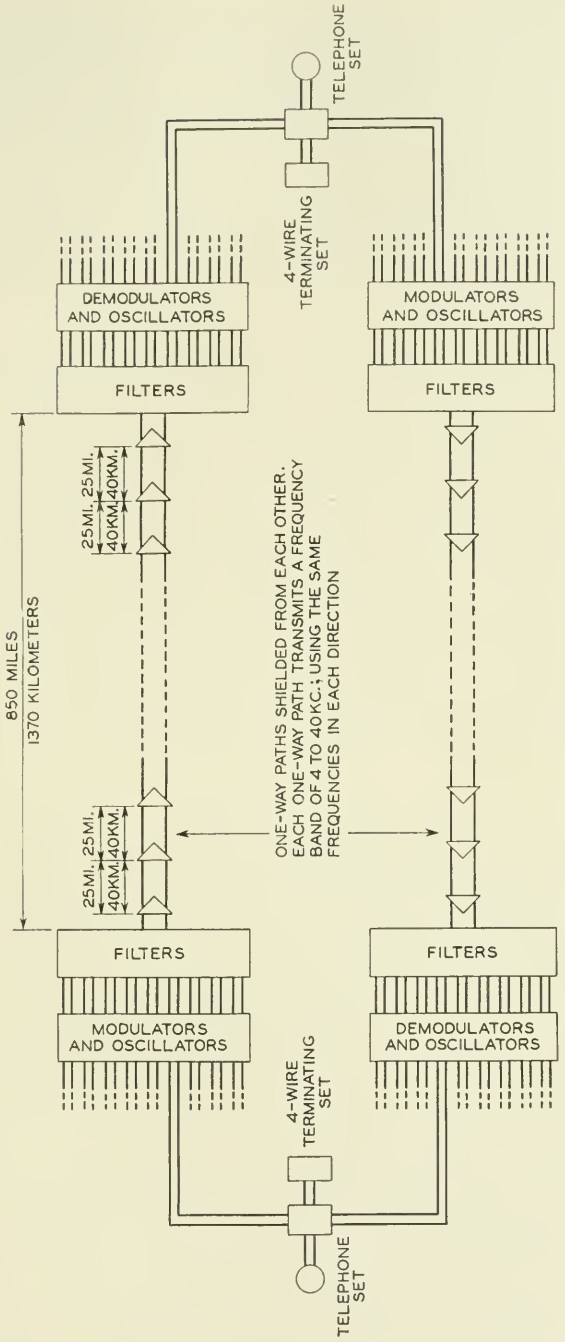

TELEPHONE SET
4-WIRE TERMINATING SET
DEMODULATORS AND OSCILLATORS
FILTERS
MODULATORS AND OSCILLATORS
FILTERS
850 MILES
1370 KILOMETERS
25 MI.
25 MI.
40 KM.
40 KM.
ONE-WAY PATHS SHIELDED FROM EACH OTHER. EACH ONE-WAY PATH TRANSMITS A FREQUENCY BAND OF 4 TO 40 KC.; USING THE SAME FREQUENCIES IN EACH DIRECTION
25 MI.
25 MI.
40 KM.
40 KM.
FILTERS
MODULATORS AND OSCILLATORS
FILTERS
DEMODULATORS AND OSCILLATORS
4-WIRE TERMINATING SET
TELEPHONE SET

Fig. 7—Experimental cable carrier telephone circuit.

nected in tandem giving an overall length of 7500 miles (12,000 kilometers) of circuit. Conversations over this 7500-mile (12,000-kilometer) circuit were very satisfactory. In fact, the transmission quality was not greatly impaired even when a 15,000-mile (24,000-kilometer) length of one-way circuit was established by connecting all of the links in tandem.

While the development of this carrier system is far from completion and it is not clear at the present time how far it can be applied to other than heavy traffic routes, it is certain that wherever this form of construction is justified, distance no longer remains as a limiting factor.

In the fifty years since the first International Electrical Congress at Paris, the new art of telephone communication has passed through many stages of development and during the past thirty years a new art, making possible communication in cable over long distance, has been born and brought to maturity. This has been made possible by a number of important and fundamental developments such as loading, quadded cable and telephone repeaters. While the development has been rapid, particularly during the past twenty years, it is not too much to expect that the next twenty or thirty years will witness an even greater and more rapid technical development and expansion in the use of long distance toll cables in all parts of the world, associated with a continued increase in the service rendered to mankind by long distance telephone communication.

Bibliography

Blackwell, O. B. The Time Factor in Telephone Transmission. *Bell System Technical Journal*, Vol. XI, January, 1932, pp. 53–66.

Carson, J. R. Building Up of Sinusoidal Currents on Long Periodically Loaded Lines. *Bell System Technical Journal*, Vol. III, October, 1924, pp. 558–566.

Clark, A. B. Telephone Transmission Over Long Cable Circuits. *Transactions of American Institute of Electrical Engineers*, Vol. 42, February, 1923, pp. 86–97. *Electrical Communication*, February, 1923, pp. 26–40. *Bell System Technical Journal*, Vol. II, January, 1923, pp. 67–94.

Clark, A. B. and C. W. Green. Long Distance Cable Circuit for Program Transmission. *Bell System Technical Journal*, Vol. IX, July, 1930, pp. 567–594.

Clark, A. B. and R. C. Mathes. Echo Suppressors for Long Telephone Circuits. *Transactions of American Institute of Electrical Engineers*, Vol. 44, April, 1925, pp. 481–490.

Crisson, G. Irregularities in Loaded Telephone Circuits. *Bell System Technical Journal*, Vol. IV, October, 1925, pp. 561-585. *Electrical Communication*, Vol. 4, October, 1925, pp. 98–110.

Demarest, C. S. Telephone Equipment for Long Cable Circuits. *Transactions of American Institute of Electrical Engineers*, Vol. 42, June, 1923, pp. 742–752. *Bell System Technical Journal*, Vol. II, July, 1923, pp. 112–140.

Lane, C. E. Phase Distortion in Telephone Apparatus. *Bell System Technical Journal*, Vol. IX, July, 1930, pp. 493–521.

Nyquist, H. and S. Brand. Measurement of Phase Distortion. *Bell System Technical Journal*, Vol. IX, July, 1930, pp. 522–549.

Osborne, H. S. A General Switching Plan for Telephone Toll Service. *Bell System Technical Journal*, Vol. IX, July, 1930, pp. 429–447.

SHACKLETON, W. J. AND I. G. BARBER. Compressed Powdered Permalloy—Manufacture and Magnetic Properties. *Transactions of American Institute of Electrical Engineers*, Vol. 47, No. 2, April, 1928, pp. 429–436.

SHAW, THOMAS AND WILLIAM FONDILLER. Development and Application of Loading for Telephone Circuits. *Bell System Technical Journal*, Vol. V, 1926 pp. 221–281. *Transactions of American Institute of Electrical Engineers*, Vol. 45, February, 1926, pp. 268–292.

STEINBERG, JOHN C. Effects of Phase Distortion on Telephone Quality. *Bell System Technical Journal*, Vol. IX, July, 1930, pp. 550–566.

ZOBEL, O. J. Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks. *Bell System Technical Journal*, Vol. VII, July, 1928, pp. 438–534.

The Conception and Demonstration of Electron Waves *

By C. J. DAVISSON

An attempt is made in this article to trace the growth of our ideas regarding the electron from their inception less than a hundred years ago to the present day. The discussion begins with a consideration of the vague and tentative deductions concerning an ultimate electrical charge which became possible when Faraday revealed the laws of electrolytic conduction; it touches upon the clarification of the conception of the electron as a charged particle capable of independent existence and subject to the laws of classical electrodynamics which was effected at the close of the last century and the beginning of the present one by the researches of J. J. Thomson and others; it indicates the difficulties in which this conception became involved, and the attempts made by Planck, Bohr and others to extricate it from them. The latter part of the paper is devoted to the amplified conception of the electron which has been developed during the last decade—a conception in which electrons are recognized as having, in different circumstances, the properties of both waves and particles.

INTRODUCTION

IT is my purpose in this report to describe a few experiments, typical of several hundred now recorded, in which streams of electrons exhibit the properties of beams of waves. It seems desirable, however, to begin by briefly reviewing various steps in the development of our conception of the electron before about the year 1925. It is against this background only that the phenomena revealed by the experiments to be described appear in true relief.

The idea that electric charge is granular was not new at the time of the first International Electrical Congress in 1881. Faraday had determined and announced the laws of electrolytic conduction fifty years earlier, and it was recognized, by some at least, that these laws suggested the existence of an elementary charge or atom of electricity. An estimate of the magnitude of this natural and presumably ultimate unit of charge had indeed been made a few years prior to the Congress by the Irish physicist Stoney from such data as were then available. The word "electron" to designate the hypothetical atom of electricity was not, however, introduced until the year 1891. The concept of the electron gained rapidly in sharpness in the decade next following—the last of the century—not so much indeed from the introduction of new ideas concerning it as from experimental evidence in support of ideas already held.

It was no new idea, for example, that neutral atoms contain positive and negative charge in equal amounts, and that the ions of an electrolyte are merely atoms or groups of atoms in which these charges are

---

* Presented by title at The International Electrical Congress, Paris, France, July 5–12, 1932.

unbalanced in one direction or the other by one or more electronic units. Yet this remained largely a speculation until the study of the conductivity imparted to gases by X-rays made all other views untenable. Ions of both signs are formed at a uniform rate within the body of a gas subjected to this then newly discovered radiation; their charges are ionic; they move through the gas by diffusion and under the influence of an impressed electric field; they disappear through recombination.

Other ideas now familiar were not entirely novel even in 1890; for instance, the idea that positive and negative electrons possess mass as well as charge—that those of one sign are more massive than the other—that within the atom the lighter revolve about the heavier ones. Weber, following on Ampere, had pictured a mechanism of this kind to explain the magnetic properties of materials. All these notions became much more plausible, however, when Lorentz showed (as he did in 1897) that the splitting and polarization of spectral lines by a magnetic field might be explained as the effect of the field upon the period of revolving particles such as Weber had assumed, and that from the magnitude of this so-called "Zeeman effect" one might actually calculate the ratio of the charge of the particle to its mass. The value so found was greater by a factor 2000, or thereabouts, than the similar ratio for hydrogen ions in electrolysis. If the charge of the particle were indeed the electronic charge, then the mass of the particle was about 1/2000 only of the mass of the hydrogen atom—a highly acceptable conclusion. The concept of the electron had gained much in definiteness, and so also had that of the atom.

But more illuminating still was the discovery made in the same year that the trajectories followed by cathode rays in traversing electric and magnetic fields are exactly those to be expected if these rays are streams of swiftly moving negatively charged particles with a charge to mass ratio amounting again, as in the foregoing instance, to about 1/2000 that of the hydrogen ion. There could be little doubt that this was the very particle inferred by Lorentz from the "Zeeman effect." Cathode rays were certainly streams of free negative electrons—unattached to atoms. This supremely important discovery was made by J. J. Thomson in England and by Wiechert in Germany.

The conception of the negative electron as a subatomic particle possessing mass as well as charge, capable of independent existence, and subject to the laws of classical electrodynamics now seemed clearly established. If various of the simple relationships were at this time sensed rather than demonstrated—such, for example, as the exact identity of the charge to mass ratios of the Zeeman effect particle and the cathode particle—there was, nevertheless, full confidence that these relationships would be confirmed by more exact measurements. And this indeed proved to be true. The anticipated details of the picture as then blocked in have since been supplied by a series of precision experiments in which Millikan's measurement of the absolute magnitude of the electronic charge is preeminent.

The turn of the century was a time of high hope. The key had been found, it appeared, to an understanding of vast ranges of phenomena; given the electron, electrodynamics and sufficient mathematics, *all* electrical and magnetic phenomena must become explicable. It seemed not too daring even to have thoughts concerning the structure of the atom. But this, as it turned out, was mostly illusion; every success of the electron theory of this period was matched by an equally conspicuous failure. Metallic conductors were pictured as containing atmospheres of free electrons with the properties of a monatomic gas. The drift of this electronic gas under the influence of an impressed field constituted the electric current. The form of Ohm's law was neatly explained, but not so the direct proportionality between the resistivity of a pure metal and its absolute temperature. The thermionic emission of electrons could be explained, apparently, in all its details, but the distribution of energy in the black body spectrum could not. The explanation of the simple Zeeman effect was most gratifying and reassuring, but the simple numerical relationships among the frequencies of line spectra remained as baffling as ever—and this, in spite of the considerable success which Drude and others had achieved in explaining the optical properties of materials in terms of electrons elastically bound within atoms.

The impasse was finally breached by Planck who showed, in 1905, that the black body spectrum could be explained if one were willing to assume that materials contain electric oscillators which emit and absorb energy only in amounts proportional to their frequencies. The conception of the electron was unaltered—not even involved, perhaps—but an oscillator, which might be a vibrating electron, was conceived to behave in a manner contrary to electrodynamical principles. A success had been achieved at the cost of violence to classical ideas regarding the production of electromagnetic radiation.

The next assault—a brilliant tour de force by Bohr—achieved its first objectives at a stride, but at a sacrifice of electrodynamical principles greater even than Planck's. Bohr showed in 1911 that by combining the idea of a concentrated atom nucleus required by Rutherford's experiments on the scattering of alpha rays with the heterodox idea of Planck, and with new devices of his own invention, one could conceive an atom model capable of yielding precisely the Rydberg constant and the complete spectrum of atomic hydrogen. The casualties included two properties previously allotted to the electron as a matter of course: the property of radiating energy during orbital motion, and the property of revolving about the nucleus in an orbit consonant with classical dynamics and determined by initial conditions which might be regarded as arbitrary. Bohr excluded from the infinity of such orbits all but a special series. The motion of the electron remained planetary, but all else was new and bizarre.

A great initial success had, however, been attained and hope of a thorough conquest of spectra ran high,—too high as it now appears, for beyond a few other quantitative successes, further achievements were qualitative to a greater or less extent and consequently less impressive. It turned out also that advancement in the elucidation of spectra could be made only at the cost of an ever increasing array of special rules and prohibitions—additional equipment of the same arbitrary nature as that of Bohr's original postulates. Out of this necessity appeared the one new idea regarding the electron which had emerged in twenty years—the idea advanced by Goudsmit and Uhlenbeck that the electron spins and possesses in consequence a magnetic moment.

It was recognized a decade ago by Bohr and others that the attack upon the atom, despite its propitious beginning, had in a considerable measure failed; and this because it had lacked, so to speak, a proper base of operation. It was felt that the many arbitrary rules and restrictions required to correlate the data of spectroscopy must flow in a natural and unforced way from fundamental mechanical principles as yet undiscovered. A system of mechanics was envisaged which would degenerate to classical mechanics for large scale phenomena, but which would present an entirely different aspect on the atomic scale, and be capable, of course, of explaining atom dynamics as revealed by spectra.

Attempts to discover these underlying principles led to the formulation by Heisenberg in 1925 of what is known as matrix-mechanics, and led L. de Broglie in the same year to put forward his ideas concerning a so-called wave-mechanics. These proposals are said to be statements in different forms of one and the same principle—so far, at any rate, as applications to atom dynamics are concerned. It is the wave-mechanics, however, which has appealed most strongly to the physicist, and it is with this only that I will here concern myself.

The basic idea of the wave-mechanics was supplied by a paradoxical situation which had existed for some years in the theory of optics. It was well established experimentally that a beam of monochromatic light can impart to individual electrons in matter amounts of energy proportional to its frequency. The factor of proportionality between these quantities—the energy imparted to the electrons and the frequency of the light—is the same as that obtained from the black body spectrum for the factor of proportionality between the energy quantum of the Planck oscillator and its frequency. The relation between the energy ϵ imparted to the electron and the frequency ν of the light is expressed, that is, by the formula $\epsilon = h\nu$, where h is the so-called Planck constant. When one tries to visualize the mechanism back of this phenomenon, he is led inevitably to a corpuscular theory of light. No other view appears adequate to explain this central fact of photoelectricity and others related to it.

On the other hand, the phenomena of interference and diffraction disposed long ago, as is well known, of an earlier corpuscular theory of light in favor of the wave theory. The demands of these phenomena are as insistent today as every they were, so that the situation comes to this, that one class of optical phenomena indicates clearly that light is a corpuscular radiation, and another indicates no less clearly that it is a propagation of waves. It is hopeless to try explaining photoelectric phenomena in terms of nothing but waves, and it is equally hopeless trying to devise a purely corpuscular interpretation of interference and diffraction.

It was de Broglie's brilliant idea that a situation similar to this might exist in regard to electrons, that electron streams like beams of light might possess in different circumstances the properties both of wave trains and of particle streams. If this were true, and if the wave aspect alone were adequate to explain the behavior of electrons in atoms, then the unhappy state of affairs which existed in regard to the interpretation of spectroscopic data might be remedied.

The formula $\epsilon = h\nu$ expresses, as we have seen, a certain correlation between the corpuscular and the wave properties of light; if the light regarded as a beam of waves is of frequency ν, then when it is regarded as a stream of corpuscles, the corpuscles or photons are of energy $\epsilon = h\nu$. A second correlation follows at once from this one and from the relation which is known to exist between the transfer of energy and of momentum by a beam of light. This second correlation relates the momentum p of the photon to the wave-length λ in vacuo of the associated waves, and is expressed by the formula $p = h/\lambda$. Or if we write σ to represent wave number—the number of waves per cm.—then the two correlations are expressed by the symmetrical formulæ

$$\epsilon = h\nu,$$
$$p = h\sigma.$$

In developing his idea of a possible wave aspect of the electron, de Broglie was led to the conclusion (partly, it appears by intuition and partly by considerations based on relativistic mechanics) that if electrons possess wave properties, the correlation between their wave and corpuscular aspects will be expressed by these same two formulæ.

It would lead us too far afield to follow even cursorily the further development of de Broglie's idea toward its original objective of explaining the behavior of atoms as disclosed by their spectra. These interesting matters must be left with the mere statement that the mathematical researches of Schrödinger and others have led to a conception of the atom in which standing wave patterns replace the permitted electron orbits of the Bohr model, and from which it is possible to derive certain laws of spectra which previously could be given only as empirical rules.

The immediate object which de Broglie had in view in postulating a wave aspect of the electron has thus been attained, and its attainment argues strongly, of course, for the soundness of the underlying conception. It is not this spectroscopic evidence, however, which reveals most clearly the wave as a real and actual property of electrons, but the more direct and unequivocal evidence supplied by experiments described in following paragraphs in which streams of electrons are diffracted by crystals.

It was implicit in de Broglie's earliest writings regarding electron waves that a stream of electrons moving with uniform speed along parallel lines will exhibit in appropriate circumstances the properties of a beam of monochromatic waves. De Broglie's first step was, indeed, to associate a train of plane parallel waves with an electron moving with uniform speed along a straight line. It remained, however, for the young German physicist Elsasser to point out the logical conclusion to which these speculations lead, and to indicate the crucial experiment by which they might be tested: to wit, that a beam of electrons scattered by an appropriate grating should exhibit the phenomenon of diffraction, and that the appropriate grating for this purpose is a crystal, since the wave-lengths calculated from de Broglie's formula for electrons of moderate speeds are like those of X-rays, of the order of one Angstrom unit.

It is the demonstration of this phenomenon—the diffraction of electrons by crystals—which constitutes now, as has been intimated, the chief experimental evidence in support of de Broglie's conception, and it is with demonstrations of this kind that I am here primarily concerned. The first of these was made by Davisson and Germer, who showed in 1927 that beams of electrons are diffracted by a crystal of nickel, and that the wave-lenghts λ deduced from the diffraction-patterns for beams of electrons of various speeds v agree with those calculated from de Broglie's formula $\lambda = h/p = h/mv$. A second and independent demonstration was made by G. P. Thomson, who showed later in the same year that beams of high speed electrons are diffracted on transmission through thin films of polycrystalline metal, and that electron wave-lengths computed from patterns so obtained verify the de Broglie relationship.

It will be well, before considering these earliest experiments in more detail, to present certain others, made more recently, of which the interpretations are more simple. The simplest experimental result which suggests that electrons should be regarded as waves rather than as particles, is perhaps the regular reflection of a beam of electrons from the face of a crystal. The experimental arrangement used in demonstrating this phenomenon is indicated on the left in Fig. 1.

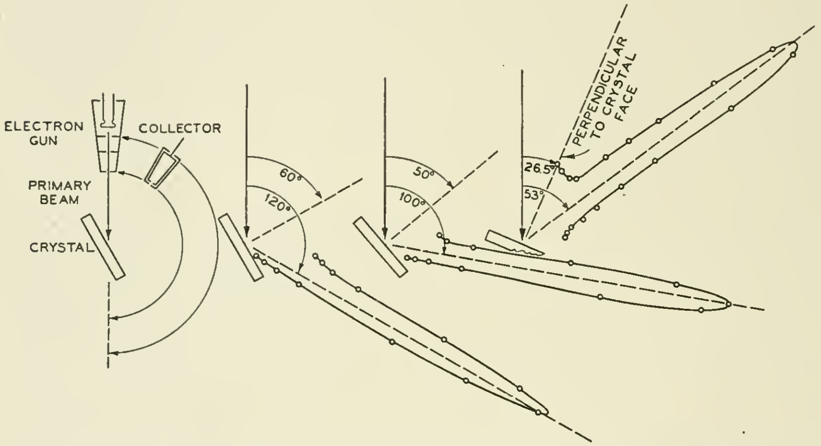


Fig. 1—Schematic diagram of apparatus for determining angular distribution of electrons scattered without loss of energy by a metallic crystal, and curves revealing specular reflection of 20-volt electrons.

Electrons emitted by a hot filament are accelerated and formed into a beam, and this beam is directed against the face of a crystal target at a known angle of incidence. The target in this case is of nickel and its face is parallel to one of the principal sets of atom planes of the crystal (111 planes). The surface is etched and presents to the incident beam a multitude of crystal facets parallel to the plane of the target face. Some of the electrons on striking these facets are scattered without loss of energy. The distribution of these full speed electrons in and near the plane of incidence is then determined by explorations with a

Faraday collector as indicated. The results of investigations of this kind, exhibited by polar diagrams in the same figure, show clearly that the incident beam is regularly reflected as if from the crystal facets.

The difficulty in explaining this result by the simple concept of electrons as particles is, that the surface of the crystal is much too coarse grained to serve as a reflector for particles as small as electrons; the diameter of the electron is of the order 10^{-13} cm., whereas, the diameters of atoms are of the order 10^{-8} cm.—greater by a factor 10^5—and this also is the order of the distance of least separation of atoms in the crystal face. It is hard to imagine how such a surface can appear smooth to the incident electrons. On the older views regarding interactions between electrons and atoms, the fate of an incident electron should be much the same as the fate of a comet plunging into a region densely packed with solar systems; the electron *might* emerge from the crystal without loss of energy after a fortunate encounter with a single atom, but its direction of departure should be a matter of private treaty between the individual electron and the individual atom—in particular, it should not be influenced by the arrangement of atoms in the crystal. The fact that nearly all of the full speed scattered electrons move away in the direction of regular reflection from the crystal face means that three atoms at least are involved in the action, since this number is required to fix the plane of the reflecting surface. The simple observation described above is thus inexplicable in terms of atoms and electrons and their interactions as previously conceived.

It is interesting to try to imagine how this phenomenon would have been interpreted, had it been observed ten years ago. With de Broglie's speculations before us, we recognize it at once as one of the circumstances in which the electron streams exhibits the properties of a wave train. The only restriction to this interpretation is that the wave-length must be assumed small compared to the linear dimensions of the reflecting surface. If wave-lengths are given correctly by de Broglie's formula they are, as has been mentioned, of the same order as those of X-rays. This explains why reflections such as exhibited in Fig. 1 are obtained from the face of a crystal, but not from a polycrystalline surface, however highly polished; the reflected beam is synthesized, so to speak from a multitude of scattered waves spreading out from atoms regularly arranged in layers parallel to the surface—lacking this regularity the synthesis does not occur.

The specular reflection of X-rays from a crystal face is, as we know, selective in the following respect: the intensity of the reflected beam passes through sharp maxima as the glancing angle θ passes through values which satisfy the Bragg formula $\sin \theta = n\lambda/2d$, where λ represents wave-length, d the distance between atom planes parallel to the surface, and n an integer. We expect this to occur with reflection of electrons, if the de Broglie waves are scattered by underlying atom layers as well as by the outermost. I will show later that the reflection portrayed in Fig. 1 is, indeed, selective. But with low-speed electrons, there are confusing complications which we can avoid by dealing with electrons of considerably higher speed; I will, therefore, begin by displaying the selective reflection of electrons accelerated through thousands rather than tens of volts only.

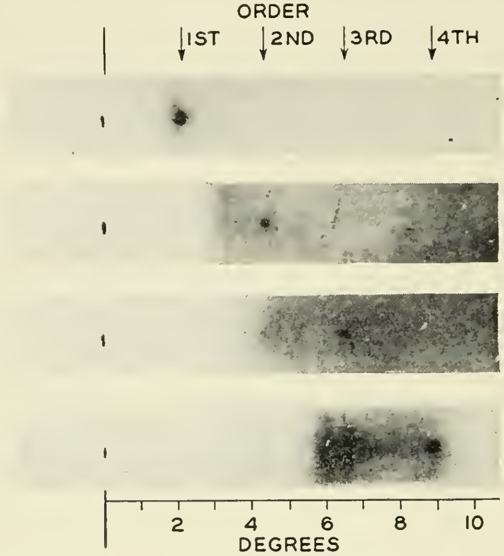


Fig. 2—Photographic record of selective reflection of 53-kv. electrons by a crystal of iron—(100) face.

In Fig. 2 we have a photographic record of the reflection of a beam of 53 kilovolt electrons from a crystal of iron. The experimental arrangement is essentially the same as that indicated in Fig. 1, though devices rather more elaborate are required for producing electron beams of such speeds. A photographic plate set at right angles to the direction of the primary beam and 30 or 40 cms. beyond the crystal replaces the exploring electrode. Each strip in Fig. 2 is the record of reflection at a particular angle of glancing. After each exposure the crystal is turned through the position of grazing (zero glancing angle), the primary beam falls directly upon the plate, and its direction is thus recorded. These fiducial marks appear in a column on the left. The glancing angles are proportional (since all of them are quite small) to the distances from the spots formed by the primary beam to the edges of the respective fogged regions; the fogging is produced by general or random scattering and its sharp cutoff on the left marks the intersection of the plane of the crystal with the photographic plate. If reflections were specular but non-selective, a spot due to the reflected beam would appear on each strip as far to the right of the fogging edge as the fiducial spot is to the left. This is not what is observed; there is strong specular reflection at a series of equally spaced angles, and at adjacent angles weaker reflection which apparently is not regular. In other ranges of angle, there is no reflection at all. This is exactly the phenomenon observed with X-rays, the apparently non-regular reflection is ascribed in the latter case to regular and selective reflection from parts of the crystal which are displaced somewhat from the mean orientation of the crystal as a whole—ascribed, that is, to imperfections in the crystal. The same explanation applies here.

Here then is an experiment in which a stream of electrons exhibits the properties of a beam of waves. The mere occurence of specular reflection is, as we have seen, incompatible with the idea that electrons are simple particles, with such properties as are commonly ascribed to particles. The further observation that the reflection is selective in accordance with the Bragg law amounts to a demonstration that we are dealing with trains of waves—or, at least, to a demonstration of the convenience of this conception—for the Bragg law is simply and accurately explained as a consequence of interference among scattered waves expanding from regularly disposed centers, such as the atoms of a crystal.

The data of the experiment enable us to calculate the length of the waves. From the Bragg law $\lambda = 2d \sin\theta/n$; the reflections are from the (100) atom planes of iron for which $d = 1.43 \times 10^{-8}$ cm. or 1.43 Ängstrom units; the value of $\sin\theta/n$ deduced from Fig. 2 and related data is 0.0189, so that $\lambda = 2 \times 1.43 \times 0.0189 = 0.054$ Ängstrom units. *This is the experimentally determined wave-length of 53 kv. electrons.*

We compare this with the *theoretical* wave-length computed from the de Broglie formula, $\lambda = h/p$. The momentum p of a particle of rest-mass m and charge e which has been accelerated from rest through a potential difference V in absolute units is given in relativistic mechanics by

$$p = (2Vem)^{1/2}\left[1 + \frac{Ve}{2mc^2}\right]^{1/2},$$

where c represents the velocity of light. Writing this into the de Broglie formula and evaluating constants for the electron one obtains as a close approximation.

$$\lambda = \left(\frac{150}{V}\right)^{1/2}[1 - 4.9 \times 10^{-7}V]$$

for V in volts. Thus the theoretical wave-length of 150-volt electrons is one Ångstrom unit or 10^{-8} cm., and the wave-length of the 53 kv. electrons employed in this experiment is 0.0546 Ångstroms which is in good agreement with the value found experimentally.

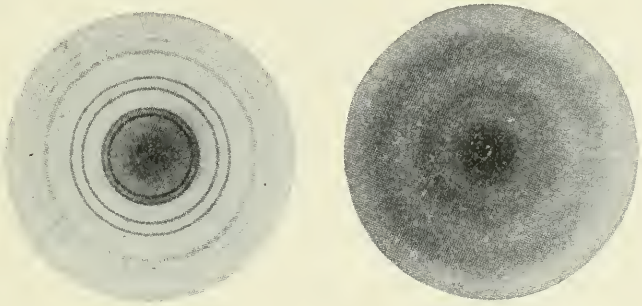
These results which are from previously unpublished data by Davisson and Germer constitute perhaps the simplest demonstration of the wave aspect of electrons and verification of the de Broglie relation—the simplest, at any rate, in which use is made of crystal diffraction.

E. Rupp has demonstrated the diffraction of electrons by an ordinary ruled optical grating and has obtained thereby values of electron wave-lengths which agree with the theoretical values within the rather wide limits of error of this border line experiment. This is, in fact, an experiment more immediately intelligible than any involving the use of crystals. But Rupp's photographic plates exhibiting these results are not very impressive, and, therefore, I have not arranged for their reproduction in this report.

The Bragg reflection, though the easiest to interpret among the types of crystal diffraction, is not the most easily demonstrated, nor the most striking, nor yet the type which yields most information concerning the diffracting crystal and the atoms composing it. It is the Hull-Debye-Scherrer type of diffraction which is in these respects preeminent, and the type which has been most thoroughly investigated. A mass of finely divided crystals of random orientation is placed in the path of a beam of monochromatic radiation; a photographic plate set at right angles to the incident beam receives and records the radiation from the diffracting material. It is evident perhaps that the pattern produced by an aggregate of a very great number of crystals oriented at random will be the same as that generated by a single crystal turned into equally many random positions. Thus, if the material is iron, the single crystal in a particular orientation gives rise to the first order diffraction spot of Fig. 2; rotate the crystal about the primary beam and this spot generates a circle or ring—so also for the second and higher spots, each of them generates a ring.

But the atoms comprising the crystal may be regarded as arranged in many different sets of planes—there are, in fact, an infinite number of such arrangements. Of these, one is unique in having a greater spacing between planes than any other—a greater value of the constant d. The first order ring due to these planes is the smallest in the pattern; the first order rings due to other arrangements follow in the order of decreasing values of d. These first order rings plus their companions of higher orders constitute the complete pattern of the aggregate. From the sequence of ring diameters, one infers the arrangement of the atoms in the crystal; from the scale of the pattern, the ratio of crystal spacing or "constant" to radiation wave-length; from the relative intensities of the rings, the way in which the scattering power of the atom varies with angle. The ring pattern is thus a storehouse of information concerning both crystal and wave.

Patterns of this type for electrons were obtained first by G. P. Thomson; it was thus, in fact, that Thomson made his demonstration of electron waves. A great many such patterns have since been obtained and studied. Two beautiful examples by Wierl are reproduced in Fig. 3. The one on the left records the diffraction of a beam of 36 kv. electrons by a thin film of polycrystalline silver, the one on the right is for 45 kv. electrons diffracted by carbon tetrachloride vapor. The form of the pattern for silver—the particular sequence of ring diameters which it displays—is characteristic of the so-called face centered cubic arrangement of atoms, and agrees with the pattern similarly obtained with X-rays. The scale of the pattern with other data of the experiment, including the scale factor of the silver crystal, yields an "observed" wave-length of 36 kv. electrons in agreement with the value computed from the de Broglie formula.



Ag CCl_4

Fig. 3—On left—diffraction pattern produced by transmission of 45-kv. electrons through thin silver foil—by R. Wierl. On right—pattern obtained by transmission of 36-kv. electrons through CCl_4 vapor—by R. Wierl.

The pattern for carbon tetrachloride is bracketed with that for silver because the chlorine atoms in this molecule have the same arrangement as the atoms in the crystal of silver. Four atoms only are required to fix a simple crystal arrangement; the chlorine atoms in CCl_4 occupy the corners of a regular tetrahedron, and this figure determines the face centered cubic arrangement. The chlorine atoms, to which nearly the whole of the scattering is due, may thus be properly thought of as forming exceedingly small crystals of this structure; the vapor is thus a crystal aggregate similar to that of polycrystalline silver. The marked differences between the two patterns are due to the disparity in the number of atoms forming the individual crystals in the two cases. The number in the chlorine crystals—four each—is the smallest possible, and the "resolving power" of the crystal, regarded as an optical instrument, is in consequence the least possible. The form of the pattern for CCl_4 accords with the tetrahedral arrangement of the chlorine atoms; the scale of the pattern with the calculated wave-length of the electrons fixes the length of the tetrahedron edge.

The purpose of this report is accomplished with the description of these few representative experiments which reveal and demonstrate—so far as demonstration is possible—a wave aspect of electrons in conformity with de Broglie's conception. The experiments do not tell us in what medium, if any, the waves occur, with what speed they are propagated, whether they are longitudinal or transverse and capable of polarization—they tell us merely that when electrons reach an element of space from a given source simultaneously over different paths the resultant intensity (the number of electrons traversing the element per unit time, as we think of it) is not the arithmetic sum of various scalar contributions as we naturally expect it to be, but is given instead by the square of the sum of contributions which are vector quantities—in precisely the manner with which we are familiar in optics. We must make the addition as if we were dealing with superposed trains of waves—with due regard for phase as well as amplitude. This kind of addition is characteristic of trains of waves. When we find quantities which add together in this particular way we conclude that the quantities are, in fact, trains of waves; this is our reason for regarding light and X-rays as wave phenomena and it is our reason also for so regarding electrons.

Having demonstrated the convenience, if not indeed the necessity, of regarding electrons in certain circumstances as waves rather than as particles, we enquire naturally if these waves are refracted on passing from one medium to another like the waves of light and X-rays, and whether or not they are polarizable. We rather expect to find that they do exhibit refraction, for if the wave-length of a beam of electrons in vacuo is given by $\lambda = (150/V)^{1/2}$ we expect that on passing into a metal of which the thermionic work function is φ the wave-length of the beam will be altered to $\lambda' = [150/(V + \varphi)]^{1/2}$. We expect, in other words, that the metal will have for electrons accelerated through V volts a refractive index given by $\mu = \lambda/\lambda' = (1 + \varphi/V)^{1/2}$. Something of this kind is indeed found experimentally, though the phenomenon is less simple than we have here imagined.

Evidence of refraction is contained in the experimental results displayed in Fig. 4. The ordinates of this curve are proportional to the intensity of the beam regularly reflected at 10 degrees incidence from a nickel crystal; the abscissa are proportional to the reciprocal of the wave-length of the incident beam. The maxima in the curve represent selective Bragg reflections of a sequence of orders, but are displaced somewhat to the left from the positions calculated from the simple Bragg formula, and indicated in the figure by arrows.

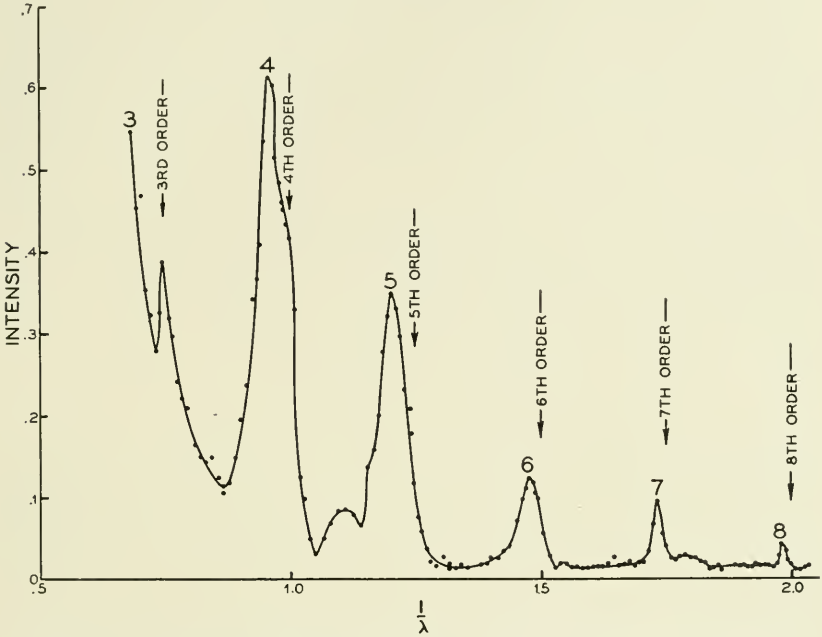


Fig. 4—Curve exhibiting selective reflection of electrons incident at 10 degrees on nickel crystal—(111) face. Departures from simple Bragg law reveal refraction.

The simple Bragg formula is derived with the assumption, however, that the refractive index of the crystal is unity. The more general formula for reflection from a crystal face is $n\lambda = 2d(\mu^2 - \cos^2 \theta)^{1/2}$ which reduces to the familiar form when the index μ is equal to one. Thus if λ_1 represents the wave-length at which Bragg reflection of a given order is expected ($\mu = 1$) and λ the wave-length at which the reflection maximum is actually observed, then the refractive index of the crystal satisfies the formula

$$(\mu^2 - 1) = \sin^2 \theta \left(\frac{\lambda^2}{\lambda_1^2} - 1 \right)$$

Or if in agreement with our assumed dispersion formula $\mu^2 - 1 = \varphi / V$ the constant φ will be given by $\varphi = (V_1 - V) \sin^2 \theta$ where V_1 and V are the voltages corresponding to wave-lengths λ_1 and λ.

On our simple view of the matter, we expect this formula to yield the same value of φ for all orders of reflection and for all angles of incidence. This expectation is not, however, realized. From Fig. 4 we obtain for the value of φ for nickel 14 or 15 volts. But under other conditions values are obtained as low as 10 volts and as high as 20 or 25 volts. The results shown in Fig. 2, if rightly interpreted, also are incompatible with the assumed law of dispersion. Thus, if θ represents the calculated glancing angle at which the nth order Bragg reflection occurs for a given wave-length when $\mu = 1$ ($\varphi = 0$), then for $\varphi \neq 0$ this reflection is to be expected at angle θ_1 such that

$$\frac{\sin \theta}{\sin \theta_1} = \left(1 - \frac{4d^2 \varphi}{150 n^2} \right)^{1/2} .$$

It will be noted that the right hand side of the equation does not involve the speed or wave-length of the incident electrons from which we conclude that relative defections from the simple Bragg law should be as conspicuous for high speed electrons as for low, and this we find not to be the case; the reflections recorded in Fig. 2 conform to the simple law, as if φ were equal to zero. The situation then is this, that while we have clear evidence that electron waves are refracted, the laws of their refraction are evidently not simple, and are yet to be discovered.

I turn now to the polarization of electron beams. Rupp, in a recent series of remarkable experiments, has shown that a beam of electrons may, in appropriate circumstances, exhibit an asymmetry with respect to its direction of motion. In these experiments, beams of electrons are diffracted by thin films of gold and annular patterns are obtained like the one for silver shown in Fig. 3. They differ, however, from the patterns ordinarily obtained in that the individual rings are not uniformly dark all around, or as we say, in all "azimuths." In a certain azimuth the density of each ring is at a maximum, in the opposite azimuth (180° away) it is at a minimum—the rings are stronger or denser on one side of the pattern than on the other. This signifies a non-uniformity or asymmetry in the number of electrons scattered into the various segments of a ring, and is due to an asymmetry in the source of the electrons constituting the beam incident upon the gold foil. These, instead of coming directly from a filament or from a high voltage discharge, as is ordinarily the case, are electrons which have already been scattered through 90 degrees by a metal target. The primary beam of electrons moved in a line parallel to the diffracting film and at some distance away; it fell upon a metal target; some of its electrons were scattered through 90° by this target, forming a secondary beam which was the one diffracted by the film of gold. Azimuths about a secondary beam formed in this way are not indistinguishable: one of them is unique in containing the primary beam which fell upon the target. It is in this azimuth that the electrons are most copiously scattered by the diffracting film, and the rings are most dense.

The sense of the effect may be stated in a different way. If the density of the rings were independent of azimuth, the mean total deflection from the primary beam (the deflection at the target plus the deflection at the film) of the electrons forming the pattern would be 90°: but actually, the mean deflection is greater than 90°. The electrons in the secondary beam have a bias toward a further deflection in the same sense as that which they received at the target.

Rupp has further shown that the effect of passing the secondary beam (the beam from the target to the film) through a *longitudinal* magnetic field is, so to speak, to rotate the azimuth of polarization; the azimuth of maximum density is displaced from its original position (without magnetic field) in a sense which depends upon the direction of the field and by an amount proportional to the length and intensity of the field.

The effect of a *transverse* magnetic field is different for different azimuths; if the direction of the field is normal, to the plane of the primary beam incident upon the target and the secondary beam the effect is nil. If it is parallel to this plane the polarization of the secondary beam decreases as the field strength is increased; at a critical intensity depolarization is complete, and at intensities which are higher still, the polarization is reversed—the azimuths of maximum and minimum density are interchanged. In these tests with transverse magnetic fields the force on the electron due to its motion through the magnetic field is balanced by the force due to a suitably adjusted electrostatic field.

The results obtained by Rupp in these experiments may be explained by means of the concepts of electron spin and corresponding magnetic moment. The spin-axes of the electrons of the primary beam we imagine to be oriented at random. But among the electrons scattered by the target this is no longer true; those scattered in any small solid angle have a non-uniform distribution in orientation; the resultant of all their spins is a vector normal to the plane containing the primary beam and the direction of scattering. Thus, the resultant spin vector is everywhere tangent to circles about the primary beam, in the same general way as the magnetic vector in the field about a wire carrying an electric current. If the polarization process is one of selection—if, that is, the scattering material selects from the incident electrons those of a particular orientation to scatter in a particular direction, then it is to be expected that double scattering will exhibit just the type of asymmetry which Rupp actually observes.

The effect of a magnetic field upon a spinning electron is to cause its spin vector to precess about an axis parallel to the direction of the field. The same statement may be made in regard to the resultant spin vector of an assemblage of electrons such as that which constitutes the beam incident upon the gold foil in Rupp's experiments. This action is competent to explain the various changes which Rupp observes in the state of polarization of the primary beam when it has traversed the differently directed fields. The magnitude of the effect yields a value of the ratio of the magnetic moment of the electron to its angular momentum and this agrees well with the theoretical value of this ratio, e/mc.

It will have been remarked perhaps that in picturing these polarization effects, we have reverted to the conception that electrons are particles. It is difficult to see how they are to be explained in terms of the characteristics of waves. There is, however, no particular reason for making the effort. Here, as in other circumstances, the choice of conception is entirely a matter of convenience.

An Expansion for Laplacian Integrals in Terms of Incomplete Gamma Functions, and Some Applications*

By EDWARD C. MOLINA

INTRODUCTION

LAPLACE has given us, in the Théorie Analytique des Probabilités, Book I, Part II, Chapter I, a method of approximating by means of series to the value of a definite integral of the type

$$I = \int_{x}^{x_2} y_1^{\theta_1} y_2^{\theta_2} \cdots y_n^{\theta_n} \phi dx,$$

where $y_1, y_2, \cdots y_n$ are functions of x whose exponents $\theta_1, \theta_2, \cdots \theta_n$ are of the same order of magnitude as a large number θ. The last function in the integrand, ϕ, embraces all factors whose exponents are of low order of magnitude compared with θ. The integral here considered must not be confused with the well known "Laplacian Transform" integral which is also embodied in the Théorie Analytique.

When the function $\phi(x)$ is other than a mere constant, the Laplacian method as presented in the Théorie Analytique does not, in certain cases, give us a series which reduces to its first term as θ approaches infinity. The object of this paper is to present a modification of the method which gives the desired result and to present two applications of considerable practical importance. The modification consists in divorcing the function ϕ from the factors of the integrand raised to high powers and associating ϕ with the factor (dx/dt) which makes its appearance with the Laplacian change of variable from x to t.

DEDUCTION OF MODIFIED EXPANSION

Setting $\theta_s/\theta = r_s$ we have

$$y_1^{\theta_1} y_2^{\theta_2} \cdots y_n^{\theta_n} = (y_1^{r_1} y_2^{r_2} \cdots y_n^{r_n})^{\theta} = y^{\theta},$$

$$I = \int_{x}^{x_2} [y(x)]^{\theta} \phi(x) dx, \tag{1}$$

which we shall write in the form

$$I = \int_{x}^{x} [y(x)^N \phi_w(x) dx,]$$

---

[*] Presented by title at International Congress of Mathematicians, Zurich, Switzerland, September, 1932.

where $N = \theta + w$, $\phi_w(x) = [y(x)]^{-w}\phi(x)$. Introducing the parameter w does not constitute an essential modification of the procedure given by Laplace, but we shall find that w plays an important part when we come to applications of the expansion presented in this paper.

In what follows it is assumed that $y(x)$ is a positive monotonically increasing or decreasing function in the range x_1, x_2; the extension to a range of integration divisible into subranges within which $y(x)$ is monotonic will be obvious.

Let X be that one of the two limits x_1, x_2 for which $y(x)$ has its greatest value; set $y(X) = Y$. Assume with Laplace that

$$\text{(2)} \qquad [\log Y - \log y(x)]^{\frac{1}{u+1}} = (x - X)g(x, X),$$

u being a positive number or zero and g a function of which $(x - X)$ is not a factor. Set

$$\text{(3)} \qquad y = Ye^{-t^{u+1}},$$

$$\text{(4)} \qquad \phi_w(x)(dx/dt) = A_0 + A_1\frac{t}{1!} + A_2\frac{t^2}{2!} + A_3\frac{t^3}{3!} + \cdots.$$

These two equations give (certain well-known conditions being fulfilled)

$$\text{(5)} \qquad I = Y^N \sum_{s=0} A_s \int_{t_1}^{t_2} \frac{t^s e^{-Nt^{u+1}} dt}{s!}\,.$$

Finally set $Nt^{u+1} = T$ and we obtain the *modified Laplacian expansion*

$$\text{(6)} \qquad I = Y^N \sum_{s=0} \left(\frac{1}{N}\right)^{H_s} B_s[P(H_s, T_2) - P(H_s, T_1)]$$

wherein

$$H_s = \frac{s + 1}{u + 1},$$

$$B_s = \left(\frac{A_s}{u + 1}\right)\frac{\Gamma(H_s)}{s!}, \qquad = A_s \quad \text{for} \quad u = 0;$$

$$T_1 = N[\log Y - \log y(x_1)], \qquad = 0 \quad \text{if} \quad X = x_1;$$

$$T_2 = N[\log Y - \log y(x_2)], \qquad = 0 \quad \text{if} \quad X = x_2;$$

$$P(H_s, T') = \frac{\displaystyle\int_0^{T'} T^{H_s - 1}e^{-T}dT}{\Gamma(H_s)},$$

$$T' = T_1 \text{ or } T_2.$$

For the computation of $P(H_s, T')$ recourse may be had to the extensive "Incomplete Gamma Function" Tables edited by Karl Pearson which give the ratio of the incomplete to the complete function. When H_s is a positive integer we have

$$P(H_s, T') = \sum_{k=H}^{\infty} \frac{T'^k e^{-T'}}{k!},$$

the well known Poisson Exponential Binomial Limit for which short tables will be found in Pearson's "Tables For Statisticians And Biometricians" and in T. C. Fry's "Probability and Its Engineering Uses."

APPLICATIONS

I

Consider the incomplete Beta Function

$$I(p) = \int_0^p (1 - x)^{n-c} x^{c-1} dx$$

for positive integral values of n and c such that n is large compared with c. Its *modified Laplacian expansion* is

$$I(p) = (c - 1)!(n + w)^{-c} \sum_{m=0} (n + w)^{-m} A(m, c - 1) P(c + m, T_2),$$

where $T_2 = -(n + w) \log (1 - p)$ and, setting $w = (d - c + 1)/2$,[1]

$$A(0, c - 1) = 1,$$

$$A(1, c - 1) = \binom{c}{1} \frac{d}{2},$$

$$A(2, c - 1) = \binom{c + 1}{2} \left[\frac{(c - 1) + 3d^2}{12} \right],$$

$$A(3, c - 1) = \binom{c + 2}{3} \left[\frac{(c - 1)d + d^3}{8} \right],$$

$$A(4, c - 1) = \binom{c + 3}{4} \left[\frac{5(c - 1)^2 - (2 - 30d^2)(c - 1) + 15d^4}{240} \right].$$

As many more coefficients as one desires can be obtained by means of the equations

$$A(m, c - 1) = \sum_{k=0}^{m} Q(m, k) \binom{c - 1 + m}{m + k},$$

$$Q(m + 1, k) = (m + k) Q(m, k - 1) + (w + k) Q(m, k),$$

$$Q(o, o) = 1.$$

[1] See Appendix II for the details of the analyses.

The expansion given above for $I(p)$ is of great practical value in connection with the evaluation of the incomplete binomial summation. We know [2] that

$$P(c; n, p) = \sum_{x=c}^{n} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \frac{n!}{(c-1)!(n-c)!} \int_0^p x^{c-1}(1-x)^{n-c} dx;$$

therefore, we have

$$P(c; n, p) = \frac{n!}{(n+w)^c (n-c)!} \sum_{m=0} \left(\frac{1}{n+w}\right)^m A(m, c-1) P(c+m, T_2).$$

This approximation formula is submitted for ranges of values of c and n such that c is of the order of magnitude of $\sqrt{n}$ or less. But for values of $P(c; n, p)$ which are of most interest, say,

$$.0001 < P < .9999,$$

c is of the same order of magnitude as

$$a = np$$

when n is a large number. Therefore, the formula is particularly valuable for values of p which are of the same order of magnitude as $1/\sqrt{n}$ or less.

The first table at the end of this paper indicates the degree of accuracy obtainable by taking $d = 0$ and using only one or two terms of the approximation for $P(c; n, p)$. The table also indicates the result of taking d such that $T_2 = a = np$, and using only one, two, three or four terms of the approximating series.

I am indebted to Miss Elizabeth McCusker and Miss Catherine Lennon of the Department of Development and Research of the American Telephone and Telegraph Company for the computation work involved in the construction of the tables at the end of this paper.

II

Consider the Bessel Function

$$I_n(b) = \frac{(\frac{1}{2}b)^n}{\sqrt{\pi}\,\Gamma(n+\frac{1}{2})} \int_0^\pi (\sin x)^{2n} e^{b(\cos x)} dx,$$

---

[2] See Laplace, "*Théorie Analytique des Probabilités,*" 1st edition, p. 151.

for b large compared with n, n being a positive integer. Writing

$$e^{-b}I_n(b) = \frac{(2b)^n}{\sqrt{\pi}\,\Gamma(n+\frac{1}{2})}\int_0^{\pi}(\tfrac{1}{2}\sin x)^{2n}\left[e^{-(\frac{1}{2}-\frac{1}{2}\cos x)}\right]^{2b}dx,$$

leads us to the *modified Laplacian expansion*

$$e^{-b}I_n(b) = \frac{\left(\dfrac{2b}{2b+w}\right)^n}{\sqrt{\pi(2b+w)}}\,P(n+\tfrac{1}{2},\,2b+w)S_n,$$

where

$$S_n = \sum_{m=0}\left(\frac{1}{8b+4w}\right)^m\frac{A_{2n+2m}}{2(2n)!}\,\frac{P(n+m+\frac{1}{2},\,2b+w)}{P(n+\frac{1}{2},\,2b+w)}\,\frac{n!}{(n+m)!}.$$ [3]

To determine the coefficients A_{2n+2m} we proceed as follows: For the integral now under consideration

$$N = 2b+w, \qquad x_1 = 0, \qquad x_2 = \pi,$$
$$y(x) = e^{-(\frac{1}{2}-\frac{1}{2}\cos x)}, \qquad X = 0, \qquad Y = 1,$$
$$(\log Y - \log y)^{\frac{1}{2}} = (\tfrac{1}{2}-\tfrac{1}{2}\cos x)^{\frac{1}{2}} = xg(x),$$
$$g(x) = \frac{1}{x}\sin\frac{x}{2} = \sum_{k=0}\frac{(-1)^k x^{2k}}{2^{2k+1}(2k+1)!},$$

so that

$$u = 1, \qquad y(x) = e^{-t^2}, \qquad T_1 = 0, \qquad T_2 = 2b+w,$$
$$\phi_w(x) = (\tfrac{1}{2}\sin x)^{2n}e^{w(xg)^2},$$
$$B_{2n+2m} = [n!\Gamma(n+\tfrac{1}{2})/(n+m)!(2n)!2^{2m+1}]A_{2n+2m}.$$

The relations above give

$$\phi_w(x)(dx/dt) = 2t^{2n}(1-t^2)^{n-\frac{1}{2}}e^{wt^2}.$$

Let

$$\sum_{s=0}A'_{2n+2s}t^{2n+2s}/(2n+2s)! = 2t^{2n}(1-t^2)^{n-\frac{1}{2}},$$

so that

$$A'_{2n} = 2(2n)!,$$
$$A'_{2n+2s} = \frac{(-1)^s(2n-1)(2n-3)\cdots(2n-2s+1)(2n+2s)!}{2^{s-1}s!};$$

then

$$\phi_w(x)(dx/dt) = e^{wt^2}\sum_{s=0}A'_{2n+2s}\frac{t^{2n+2s}}{(2n+2s)!}$$
$$= \sum_{k=0}\sum_{s=0}t^{2n+2s+2k}A'_{2n+2s}w^k/k!(2n+2s)!$$

---

[3] An expansion of this type for Bessel functions was derived many years ago by Hadamard; see Watson, "A Treatise on the Theory of Bessel Functions," pp. 204 and 205.

and

$$A_{2n+2m}/(2n+2m)! = \sum_{k=0}^{m} (w^k/k!) A'_{2n+2m-2k}/(2n+2m-2k)!$$

We must now assign a value to w. One which suggests itself is to take w such that the second term in our expansion for $I_n(b)$ vanishes; in other words, such that A_{2n+2} shall be zero. This merely requires that we set $w = n - 1/2$, giving

$A_{2n} = 2(2n)!, \qquad A_{2n+2} = 0$

$A_{2n+4} = -(2n-1)(2n+4)!/2, \qquad A_{2n+6} = -(2n-1)(2n+6)!/3$

$A_{2n+8} = (2n-1)(2n-5)(2n+8)!/16$, etc.

For $n = 1$ and $w = 1/2$ we have

$$A_2 = 4, \qquad A_4 = 0, \qquad A_6 = -360, \qquad A_8 = -13440, \cdots$$

$$e^{-b} I_1(b) = \left[\left(\frac{2b}{2b+.5} \right) P(3/2, 2b+.5)/\sqrt{\pi(2b+.5)} \right] S_1$$

wherein

$$S_1 = 1 - 15(8b+2)^{-2} \frac{P(7/2, 2b+.5)}{P(3/2, 2b+.5)} - 140(8b+2)^{-3} \frac{P(9/2, 2b+.5)}{P(3/2, 2b+.5)} - \cdots$$

The degree of accuracy obtainable when only one, two or three terms of S_1 are made use of is indicated by the second table at the end of this paper. For some values of b, ranging from 6 to 16, comparison is made between the successive approximations for $e^{-b}I_1(b)$ and its true value as given by Watson in his Theory of Bessel Functions. The last three columns of Table II give the figures for $w = 0$ instead of $w = 1/2$.

Appendix I

When the expansion of $\phi_w(x)(dx/dt)$ is not obvious we may proceed as does Laplace in expanding (dx/dt).

Equations (2) and (3) give

$$x = X + tg^{-1}. \tag{7}$$

This last equation, together with the Lagrange-Laplace expansion theorem, gives for a function of x the expansion

$$f(x) = f(X) + \sum_{s=1} \frac{t^s}{s!} \left[\frac{d^{s-1}}{dx^{s-1}} \left(\frac{1}{g} \right)^s f'(x) \right]_{x=X}.$$

But $df(x)/dt = f'(x)(dx/dt)$. Therefore, taking $f(x)$ such that $f'(x) = \phi_w(x)$,

$$\phi_w(x)\frac{dx}{dt} = \sum_{s=0}\frac{t^s}{s!}A_s,$$

$$A_s = [D_x{}^s g^{-(s+1)}\phi_w(x)]_{x=X}, \qquad D_x{}^s = \frac{d^s}{dx^s},$$

$$= Y^{-w}\sum_{v=0}\frac{w^v}{v!}[D_x{}^s(x - X)^{(u+1)v}g^{-(s+1)+(u+1)v}\phi(x)]_{x=X},$$

or

$$\frac{A_s}{s!} = Y^{-w}\sum_{v=0}\frac{w^v}{v!}\frac{[D_x{}^{s-(u+1)v}\phi(x)g^{-(s+1)+(u+1)v}]_{x=X}}{[s - (u + 1)v]!}. \tag{8}$$

Beta Function

For the incomplete Beta Function we have,

$$X = 0, \qquad u = 0, \qquad Y = 1, \qquad g(0) = 1,$$

$$\phi(x) = x^{c-1}(1 - x)^{-c} = \sum_{r=0}\binom{c - 1 + r}{r}x^{c-1+r}.$$

Therefore,

$$\frac{A_s}{s!} = \sum_{v=0}^{s-(c-1)}\frac{w^v}{v!}\sum_{r=0}\binom{c - 1 + r}{r}\frac{[D_x{}^{s-v}x^{c-1+r}g^{-(s+1-v)}]_{x=0}}{(s - v)!},$$

or

$$\frac{A_s}{s!} = \sum_{v=0}^{s-(c-1)}\frac{w^v}{v!}\sum_{r=0}^{s-(c-1)-v}\binom{c - 1 + r}{r}\frac{[D_x{}^{s-(c-1)-v-r}g^{-(s+1-v)}]_{x=0}}{[s - (c - 1) - v - r]!}.$$

Since $A_s = 0$ for $s < c - 1$, set $s = c - 1 + m$; then interchanging v and $m - v$ we obtain

$$\frac{A_{c-1+m}}{(c - 1 + m)!} = \sum_{v=0}^{m}\frac{w^{m-v}}{(m - v)!}\sum_{r=0}^{v}\binom{c - 1 + r}{c - 1}\frac{[D_x{}^{v-r}g^{-(c+v)}]_{x=0}}{(v - r)!}. \tag{9}$$

To evaluate (9) we may have recourse to the formula[4]

$$D_x{}^M g^{-S} = S\binom{S + M}{M}\sum_{K=1}^{M}\left[\frac{(-1)^K\binom{M}{K}D_x{}^M g^K}{(S + K)g^{S+K}}\right], \tag{10}$$

taking

$$M = v - r,$$
$$S = c + v.$$

---

[4] See formula 25, page 15, of Schlömilchs Höheren Analysis, Band II, Dritte Auflage, 1879. I am indebted to Mr. J. Riordan of the Department of Development and Research, American Telephone and Telegraph Company, for calling my attention to this formula.

Bessel Function

For the Bessel integral with b large compared with n we have

$$X = 0, \qquad u = 1, \qquad g(0) = 1/2, \qquad Y = 1,$$

$$\begin{aligned}\phi(x) &= \left\{\left[\sin\left(\frac{x}{2}\right)\right]^2\left[1 - \sin^2\left(\frac{x}{2}\right)\right]\right\}^n = [(xg)^2 - (xg)^4]^n \\ &= \sum_{r=0}^{n}\binom{n}{r}(-1)^r(xg)^{2n+2r}.\end{aligned}$$

Substituting in (8) we have

$$\frac{A_s}{s!} = \sum_{v=0}\frac{w^v}{v!}\sum_{r=0}^{R}(-1)^r\binom{n}{r}\frac{\left[D_x^{\,s-2v}x^{2n+2r}g^{-(s-2n-2v-2r+1)}\right]_{x=0}}{(s-2v)!},$$

where $2R \not> s - 2n - 2v$; note that if $R > n$, then $\binom{n}{r} = 0$.

Since $A_s = 0$ for $s < 2n$ and g is an even function of x, set $s = 2n + 2m$. Then Leibnitz's theorem and interchanging v with $m - v$ give

$$\text{(11)} \qquad \frac{A_{2n+2m}}{(2n+2m)!} = \sum_{v=0}^{m}\frac{w^{m-v}}{(m-v)!}\sum_{r=0}^{v}(-1)^r\binom{n}{r}\frac{\left[D_x^{\,2v-2r}g^{-(2v-2r+1)}\right]_{x=0}}{(2v-2r)!},$$

where again we may have recourse to (10) for the differentiations of negative powers of $g(x)$.

APPENDIX II

Writing the incomplete Beta Function in the form

$$I(p) = \int_0^p (1-x)^{n+w}x^{c-1}(1-x)^{-(c+w)}dx,$$

we now have

$$\begin{aligned}&N = n + w, \qquad x_1 = 0, \qquad x_2 = p, \\ &y(x) = (1 - x), \qquad Y = 1, \qquad X = 0, \\ &\phi_w(x) = x^{c-1}(1-x)^{-(c+w)}, \\ &\log Y - \log y = -\log(1-x), \\ &\qquad\qquad = xg(x), \qquad g(x) = \left(1 + \frac{x}{2} + \frac{x^2}{3}\cdots\right), \\ &u = 0, \\ &y(x) = e^{-t}, \\ &T_1 = 0, \\ &T_2 = -(n+w)\log(1-p).\end{aligned}$$

From these equations we derive

$$\begin{aligned}\phi_w(x)(dx/dt) &= e^{wt}(e^t - 1)^{c-1}\\ &= \sum_{k=0}^{c-1}(-1)^k\binom{c-1}{k}e^{(w+c-1-k)t}\\ &= \sum_{m=0}\left(\frac{t^s}{s!}\right)\left[\sum_{k=0}^{c-1}(-1)^k\binom{c-1}{k}(w+c-1-k)^s\right],\\ &\qquad\qquad s = c-1+m.\end{aligned}$$

Therefore (see any work on finite differences),

$$A_{c-1+m} = \Delta^{c-1}(w)^{c-1+m} = (-1)^m\Delta^{c-1}(-w-c+1)^{c-1+m}.$$

Now

$$\begin{aligned}\Delta^{c-1}(w)^{c-1+m+1} &= w\Delta^{c-1}(w)^{c-1+m} + (c-1)\Delta^{c-2}(w+1)^{c-1+m}\\ &= (w+c-1)\Delta^{c-1}(w)^{c-1+m} + (c-1)\Delta^{c-2}(w)^{c-2+m+1}.\end{aligned}$$

Increasing c by 1, decreasing m by 1 and setting

$$A_{c-1+m} = (c-1)!A(m, c-1),$$

we have

$$A(m, c) - A(m, c-1) = (w+c)A(m-1, c)$$

or

$$\Delta_c A(m, c-1) = (w+c)A(m-1, c)$$

where the subscript c implies that Δ_c operates on c. Assume that

$$A(m, c-1) = \sum_{k=0}^{m} Q(m, k)\binom{c-1+m}{m+k},$$

where the coefficients $Q(m, k)$ are functions of w; then

$$\begin{aligned}\Delta_c A(m+1, c-1) &= \sum_{k=0}^{m} Q(m, k)(w+c)\binom{c+m}{m+k}\\ &= \sum_{k=0}^{m+1} Q(m, k)\left[(m+k+1)\binom{c+m}{m+k+1} + (w+k)\binom{c+m}{m+k}\right]\\ &= \sum_{k=0}^{m+1}[(m+k)Q(m, k-1) + (w+k)Q(m, k)]\binom{c+m}{m+k},\end{aligned}$$

from which, by finite integration with reference to c, we obtain

$$A(m+1, c-1) = \sum_{k=0}^{m+1}[(m+k)Q(m, k-1) + (w+k)Q(m, k)]\binom{c-1+m+1}{m+1+k}$$

so that

$$\begin{aligned}Q(m+1, 0) &= wQ(m, 0),\\ Q(m+1, k) &= (m+k)Q(m, k-1) + (w+k)Q(m, k),\\ &\qquad\qquad 0 < k < m+1,\\ Q(m+1, m+1) &= (2m+1)Q(m, m).\end{aligned}$$

Since

$$A(0, c-1) = 1 = \binom{c-1}{0},$$

we know that $Q(0, 0) = 1$. Hence

$$A(1, c-1) = w\binom{c}{1} + \binom{c}{2},$$

$$A(2, c-1) = w^2\binom{c+1}{2} + (3w+1)\binom{c+1}{3} + 3\binom{c+1}{4},$$

$$A(3, c-1) = w^3\binom{c+2}{3} + (6w^2+4w+1)\binom{c+2}{4} + (15w+10)\binom{c+2}{5} + 15\binom{c+2}{6}.$$

$$A(4, c-1) = w^4\binom{c+3}{4} + (10w^3+10w^2+5w+1)\binom{c+3}{5} + (45w^2+60w+25)\binom{c+3}{6} + (105w+105)\binom{c+3}{7} + 105\binom{c+3}{8},$$

etc.

As yet the value of w has not been specified. Since when m is an odd number and $w = -(c-1)/2$, the equation

$$\Delta^{c-1}(w)^{c-1+m} = (-1)^m \Delta^{c-1}(-w-c+1)^{c-1+m}$$

gives us

$$\Delta^{c-1}\left(-\frac{c-1}{2}\right)^{c-1+m} = 0,$$

let us write $w = (d/2) - (c-1)/2$, leaving d arbitrary; we obtain

$$A(0, c-1) = 1,$$

$$A(1, c-1) = \binom{c}{1}\frac{d}{2},$$

$$A(2, c-1) = \binom{c+1}{2}\left[\frac{(c-1)+3d^2}{12}\right],$$

$$A(3, c-1) = \binom{c+2}{3}\left[\frac{(c-1)d+d^3}{8}\right],$$

$$A(4, c-1) = \binom{c+3}{4}\left[\frac{5(c-1)^2-(2-30d^2)(c-1)+15d^4}{240}\right],$$

etc. Finally, for $d = 0$, or $w = -(c-1)/2$,

$$A(0, c-1) = 1, \qquad A(1, c-1) = 0,$$

$$A(2, c-1) = \binom{c+1}{3}\frac{1}{4}, \qquad A(3, c-1) = 0,$$

$$A(4, c-1) = \binom{c+3}{5}\left(\frac{5c-7}{48}\right), \qquad A(5, c-1) = 0$$

etc.

TABLE I

Incomplete Binomial Summation

$$P(c;\,n,\,p) = \frac{n!}{(c-1)!(n-c)!}\int_0^p x^{c-1}(1-x)^{n-c}dx$$

np	n	p	c	$P(c;\,n,\,p)10^6$	$w \doteq -(c-1)/2$		w Such that $T_2 = np$			
					$(P-P_1)10^6$	$(P-P_2)10^6$	$(P-P_1)10^6$	$(P-P_2)10^6$	$(P-P_3)10^6$	$(P-P_4)10^6$
5	50	.1	2	966,214	78	0	− 77,687	4,340	− 198	8
			4	749,706	428	0	− 51,250	2,399	− 90	3
			6	383,877	518	0	− 802	520	− 2	0
			8	122,145	273	0	9,911	676	34	2
			10	24,538	77	0	4,058	426	34	2
			12	3,220	13	0	789	117	13	1
	100	.05	2	962,919	19	0	− 36,837	986	− 22	0
			4	742,161	101	0	− 23,907	535	− 10	0
			6	384,001	123	0	− 190	123	0	0
			8	127,960	68	0	5,209	175	4	0
			10	28,188	21	0	2,378	124	5	0
			12	4,274	4	0	547	40	2	0
10	100	.1	2	999,678	25	0	− 98,756	7,193	− 460	27
			4	992,164	244	0	− 155,547	14,675	− 1,085	68
			6	942,423	739	0	− 152,886	14,185	− 997	58
			8	793,949	1,268	1	− 94,502	7,232	− 430	22
			10	548,710	1,399	2	− 25,277	2,065	− 83	4
			12	296,967	1,054	2	11,472	1,244	43	3
			14	123,877	563	1	15,423	1,511	112	7
			16	39,890	221	1	8,242	1,102	112	10
			18	10,007	65	0	2,832	495	65	7
			20	1,979	15	0	699	152	24	3

TABLE I—*Continued*

np	n	p	c	$P(c; n, p)10^6$	$w = -(c-1)/2$		w Such that $T_2 = np$			
					$(P - P_1)10^6$	$(P - P_2)10^6$	$(P - P_1)10^6$	$(P - P_2)10^6$	$(P - P_3)10^6$	$(P - P_4)10^6$
	200	.05	2	999,596	6	0	− 47,020	1,644	− 51	1
			4	990,951	59	0	− 72,566	3,261	− 116	3
			6	937,657	177	0	− 70,458	3,100	− 104	3
			8	786,695	300	0	− 43,386	1,575	− 45	1
			10	545,290	330	0	− 11,460	462	− 9	0
			12	300,244	252	0	6,002	309	5	0
			14	129,892	140	0	8,241	402	15	0
			16	44,355	58	0	4,750	318	16	0
			18	12,089	18	0	1,811	160	10	0
			20	2,664	4	0	509	56	4	0

$P - P_m$ = error incurred by using only the first m terms of the series for $P(c; n, p)$ presented in this paper.

TABLE II

BESSEL FUNCTION
$e^{-b}I_1(b)$

b	P	$w = n - 1/2 = 1/2$ for $n = 1$			$w = 0$		
		$\left(\frac{P - P_1}{P}\right)$	$\left(\frac{P - P_2}{P}\right)$	$\left(\frac{P - P_3}{P}\right)$	$\left(\frac{P - P_1}{P}\right)$	$\left(\frac{P - P_2}{P}\right)$	$\left(\frac{P - P_3}{P}\right)$
6	.152052	−.007497	−.001457	−.000331	−.071134	−.004188	−.000701
8	.134143	−.004043	−.000587	−.000098	−.051474	−.002187	−.000261
10	.121263	−.002529	−.000293	−.000038	−.040358	−.001344	−.000125
12	.111464	−.001731	−.000167	−.000018	−.033198	−.000911	−.000070
14	.103698	−.001259	−.000104	−.000010	−.028198	−.000658	−.000042
16	.097350	−.000958	−.000070	−.000006	−.024510	−.000497	−.000029

P = value of $e^{-b}I_1(b)$ from Table II of Watson's Theory of Bessel Functions.

$\left(\frac{P - P_m}{P}\right)$ = proportional error incurred by using the first m terms of the series for $e^{-b}I_1(b)$ presented in this paper.

Contemporary Advances in Physics, XXIV

High-Frequency Phenomena in Gases, First Part

By KARL K. DARROW

This is an account of the behavior of conducting gases subjected to high-frequency electrostatic fields—behavior which can be interpreted, in many cases with striking success, by supposing that the free electrons wandering in the gas are set into motion by the field, and oscillate and drift according to laws which can be derived from our knowledge of the response of free electrons to steady fields. When a constant magnetic field coexists with the high-frequency forces, the phenomena become more complicated, but remain predictable. There are also peculiar phenomena indicating that the electrons in a conductive gas have certain natural frequencies of oscillation. Applications are made to the absorption of radio-frequency waves in ionized gases.

IN this article I will describe some of the phenomena which are observed when the voltage across a region filled with gas is varying quite rapidly. Considered as a function of time, the voltage may be periodic, a sine-wave with uniform amplitude and of a frequency somewhere between a few thousands and a few hundreds of millions of cycles per second. It may be a succession of highly-damped wavetrains, each commencing with a rapid rise of voltage and continuing in oscillations of high frequency but swiftly declining amplitude, which die away into nothing and after an interval (it may be of a few hundredths or a few thousandths of a second) are followed by another train. It may be a brief irregular spasm of electromotive force, of which the highest value of the voltage is measured or merely guessed. In the gas itself there may be the phenomenon of sudden and violent breakdown; or the establishment of a self-sustaining luminous discharge, like in aspect to a glow or an arc; or merely a vibratory motion of electrons, freed by other agencies and set in motion by the oscillating field.

One might regard this as a subject which physicists had better leave alone, until they have full understanding of the seemingly much easier problems of discharges across a gas exposed to a constant voltage. So great are the apparent advantages of steady or "direct-current" discharges for the student, so great the apparent inconveniencies of high frequencies, that one might reasonably think it futile to assail the latter with weapons which have not yet overcome the former. Thus, in a self-sustaining glow maintained by a constant voltage, there is a peculiar distribution of space-charge, which distorts the field between the electrodes in a characteristic way, and manifests itself by a striking subdivision of the gas into zones of light and zones of darkness, differing much from one another in their electrical state as well as in their aspect. This distribution is not established instantly; would one not do better to examine it at leisure and understand it in full for each separate value of voltage, before beginning to study discharges in which it is continually changing over from the form appropriate to one voltage into the form appropriate to another? or discharges in which the voltage is varying so rapidly, that the gas is always in a state of transition, never even approaching the equilibrium appropriate to any steady value of voltage whatever?

Well, this is not necessarily to be supposed! it may be that when the voltage across the gas varies with extreme rapidity, the gas itself enters into a sort of equilibrium-state, perhaps even a more intelligible state than that which it attains when it is allowed an ample time to adjust itself to a constant value of voltage. The simplicity of certain empirical laws of the high-frequency glow suggests this, as also does the aspect of the glow. There is also the following argument, rather paradoxical in sound, perhaps, but forcible. A self-sustaining direct-current discharge, a glow or an arc, involves a steady outflow of electrons from its cathode. This outflow must be maintained by agents coming out of the gas itself—photons and positive ions and excited atoms, which are generated in the gas by the discharge and fall upon the cathode. Of the relative prominence of these agents, little is known—it forms one of the major problems of the steady discharge in gases; but at least it is certain, that they owe their existence and their effectiveness largely to the distribution of space-charge in the gas. The distribution of the space-charge therefore is controlled by the requirement (or seems controlled—one never knows what is cause and what is effect) that the electron-outflow from the cathode must be kept at a level suitably high. Now in the high-frequency discharge, the demand for electrons from the cathode is attenuated or abolished; witness the fact that such discharges may be maintained when the electrodes are separated by insulators from the gas! The peculiar disposition of the space-charge is therefore not demanded, at least not to so great an extent; conditions are intrinsically simpler.

This feature of high-frequency discharges—their competence to do without electrons from the cathode—requires attention. It is derived from a fundamental principle, which one is all too likely to forget if one has long been occupied with steady currents: the principle of Maxwell, that an electric field when varying in time is equivalent to a current. Let us apply this principle to a current in a circuit (Fig. 1) composed of a source E of electromotive force; wires leading (one of them through a galvanometer G) from the opposite poles of the source to electrodes A and C; and a gas between the electrodes. For convenience let us imagine the gas enclosed in a cylindrical vertical tube, the electrodes near its ends.

Take up first the direct-current discharge; say the cathode is above; consider any surface, S, cutting across the gas which the tube encloses. After the steady state is established, there will be positive ions crossing the surface on their way toward the cathode (upward) and negative ions crossing it on their way toward the anode (downward). We may speak of these, both classes, as ions going in the "right" direction. The total charge which they carry, as many of them as cross in unit time, will be the current across the surface —if there are no ions going in the "wrong" directions. But in a luminous discharge there are generally positive ions crossing toward the anode and negative ions crossing toward the cathode. The sum of the charges which they carry must be subtracted from the former sum; or, to express the idea better, they must be added with a

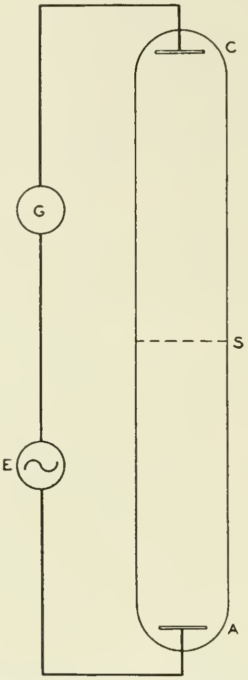


Fig. 1—Illustrating an internal-electrode tube for the transmission of high-frequency current through gas.

negative sign. We have to form a fourfold sum: charges borne by positive ions going up and charges borne by negative ions going down, with *plus* sign; charges borne by negative ions going up and charges borne by positive ions going down, with *minus* sign. The terms of this fourfold sum may vary from one to another of the surfaces intersecting the tube (in general, they do) but the sum remains the same; it is the current through the gas. It is also the current flowing through the plane of the cathode-surface (it may then consist of two terms only, electrons emerging from the cathode metal and positive ions impinging on it; or it may involve the two additional terms, because of electrons falling upon the cathode and positive ions rebounding from it). It is also the current flowing through the wires, and the current measured by the galvanometer at G.

Now for a high-frequency discharge, this must be modified. The fourfold sum aforesaid is not the whole of the current. Or rather, we might call it the whole of the current, but then we should be compelled to say that the current across the various cross-sections of the circuit is not the same, and is not measured by the galvanometer. We do better to follow Maxwell, or rather the usage which developed out of Maxwell's theory: call it the "net convection current" and introduce the name "displacement current" for the quantity which must be introduced, in order to get a sum which is the same for all the cross-sections of the circuit and equal to the reading of the galvanometer. Nor is this "sum" to be obtained by simple addition. We are obliged to take into account another complication, from which direct-current discharges are free: the necessity of distinguishing between a current-component which is proportional to the voltage and a current-component which is proportional to the rate of change of the voltage; or, to take the simple case of a sinusoidal voltage, the current-components which are respectively "in phase" and "in quadrature" therewith. The displacement-current belongs entirely to the latter component, while the convection-current may belong partly to the one and partly to the other, as I will presently illustrate.

It is, then, *not* required that the net convection-current should be the same all through the gas; much less, that it should be the same as the current reported by the galvanometer. At sufficiently high frequencies, electrons and positive ions may oscillate in the interspace between anode and cathode without reaching either, if they start their independent careers far enough off from both. Can one attain a condition in which there is a convection-current of oscillating ions in the middle of the gas, but none whatever in the vicinity of the electrodes, and no electrons escape from the metal of the cathode into the gas? Without an impervious screen between the metal and the gas this would, I suppose, be impracticable; but if there be such a screen—if for instance the electrodes are outside of the glass-walled tube containing the gas, instead of being inside—it can be achieved. The ions in the imprisoned gas can be set into oscillatory motion by an alternating voltage applied to electrodes outside; indeed, a self-sustaining luminous glow-discharge can be maintained within the tube; the charged particles which vibrate in the shining gas never enter the metal parts of the circuit nor transfer their charges to these. Only by invoking Maxwell's concept of the displacement current are we able to contend that there is continuity of current-flow around the circuit and across the gap between the anode and the cathode.

The so-called "current through the gas"—i.e. the reading of the galvanometer in series with the circuit—is therefore a datum decidedly remote from the phenomena within the gas. Even in the simplest of all cases, that of a direct-current discharge, it merely gives the sum of the four terms aforesaid, not the value of any individually. In the much more intricate case of the high-frequency discharge, it is a combination of the four terms aforesaid and the displacement-current. This is probably why it figures so little in accounts of the high-frequency discharge.

If the current through the gas is hard to interpret when measured, the voltage across it may present an easier problem, or may on the other hand be unmeasurable altogether. The simplest case of all is that of a gas in contact with electrodes (as in the tube of Fig. 1) between the two of which an undamped sinusoidal voltage is applied. An alternating-current electrostatic voltmeter, of one type or another, shunted across the electrodes, will give the "effective" or "root-mean-square" value of the potential-difference between them. It is often the maximum value of the voltage which the experimenter wants, or thinks that he wants; to get it he must multiply the reading of the voltmeter by $\sqrt{2}$. This is not the proper factor unless the voltage is truly sinusoidal; the observer must find out about this (by using an oscillograph to make the waveform visible, or by hunting with a wavemeter for harmonics of the fundamental frequency) and use a different factor if the waveform is distinctly not a pure sine-function. If the electrodes are outside of the tube containing the gas, some part of the potential-gradient between them is located in the insulating walls, and does not act upon the gas. If the voltage is applied as a sequence of highly-damped short wavetrains with intervals between, the problem of determining its maximum value is a serious one. Perhaps the best available methods have never been applied in experiments in this field; at any rate, many a physicist has contented himself with measuring something which he thought to be proportional to the maximum voltage, without knowing what value to assign to the factor of proportionality.

The toughest problem of all, in respect of measuring current and voltage—or let me say, in respect of finding something significant to measure—is forced upon us by a form of discharge which Hittorf invented (this is probably a more suitable word than "discovered") in 1884. He wound a wire spirally around a tube containing air at low pressure, connected the spiral across a Leyden jar and the terminals of an induction-coil, and thus sent through the wire a sequence of current-pulses which were highly damped wavetrains; they incited a brilliant glow in the tube. Within the spiral, and therefore pervading the gas in the tube, there was of course a magnetic field parallel to the axis and alternating its direction with the alternations of the pulsing current. There was also a circular electric field due to the varying magnetic field, pointing alternately clockwise and counterclockwise around the axis. Also there was a varying electric field due to the alternating potential-differences between the windings of the spiral. All of these three must have influenced the mobile ions of the gas! Often, as Thomson was later to stress, the discharge takes the form of a brilliant ring, thus suggesting that it is the second of the fields which dominates. But the problem of their relative responsibilities is a subtle and very difficult one; and the "ring discharge" is as troublesome to elucidate by theory, as it is easy to realize in practice.

I have written thus far as though we were concerned entirely with two sorts of stable conditions: the self-sustaining luminous high-frequency discharge, and the oscillations of ions in a gas where the supply of ions is maintained not by the alternating voltage but by some other agency acting independently. We are concerned with these, and with a third matter as well: the process of "breakdown," the sudden onset of the self-sustaining discharge which occurs when the amplitude of the high-frequency voltage across a gas hitherto tranquil is elevated past a critical value. This critical value or "breakdown-potential" is the most frequently measured of all the measurable qualities of the discharge; though strictly speaking it is scarcely a quality of the discharge, but rather of the tranquil gas of which it marks the oncoming transformation. Presumably it is preceded by an intermediate state, in which the oscillating voltage both displaces the ions in the gas, and gives them energy enough to form others; but of this we as yet know little.

I will arrange the optics much as I have previously arranged them in the description of direct-current discharges:[1] first, the observations on gases independently ionized and exposed to an oscillating voltage which displaces the independently-formed ions; then the observations on breakdown or transition, in which the oscillating voltage is raised to such a value that it initiates and maintains a self-sustaining discharge; and finally, the observations on the self-sustaining discharge itself. Beforehand, though, it will be convenient to derive some equations describing the presumable behavior of ions in gases exposed to alternating fields.

First, consider a charged particle free to move in a vacuum, and exposed to an alternating electric field. One might expect it to oscillate to and fro across a fixed point, like a simple harmonic vibrator. This however occurs only in a special case. In general, the particle will oscillate about a point which glides at uniform speed along the direction of the field. For, let us write down the equation for the acceleration of the particle, denoting by e its charge, by m its mass, by E_0 the amplitude and by ν the frequency of the field E, which last we suppose directed along the axis of x:

$$m(d^2x/dt^2) = eE = eE_0 \sin 2\pi\nu t. \tag{1}$$

Integrating once, we obtain for the speed of the ion,

$$dx/dt = \frac{1}{2\pi\nu}\frac{eE_0}{m}(1 - \cos 2\pi\nu t) + v_0, \tag{2}$$

and integrating a second time, we obtain for its displacement from the point which it occupied at $t = 0$:

$$x = \left(\frac{1}{2\pi\nu}\frac{eE_0}{m} + v_0\right)t - \frac{1}{4\pi^2\nu^2}\frac{eE_0}{m}\sin 2\pi\nu t. \tag{3}$$

The first term on the right of equation (3) represents the steady drift of the centre of oscillation, the second term the oscillation to and fro across this centre.

The speed of the steady drift is the coefficient of t in equation (3). Its value depends on that of the constant of integration v_0, which, as one sees from equation (2), is the speed (or rather, the x-component of the velocity) of the ion at the instant $t = 0$ when the electric field passes through zero. If this speed just happened to be equal to $eE_0/2\pi\nu m$, and directed in the sense opposite to that in which the increasing field was about to draw the ion, the particle would oscillate about a fixed point; the amplitude of its vibrations would amount to $eE_0/4\pi^2\nu^2 m$, its maximum kinetic energy to $e^2E_0^2/8\pi^2\nu^2 m$. But this is a very improbable event; it is difficult to conceive of any mechanism plausible or unplausible which would endow all the ions in a gas with such an initial speed. We can scarcely make use of any but the most general formulæ; among which I choose those for the drift-speed u and the maximum kinetic energy K_m of the ions:

---

[1] In a recent book, "Electrical Phenomena in Gases," abstracted in the July 1932 *Bell Sys. Tech. Jour.*, and to which I refer at various points in this article.

$$\begin{aligned} u &= \frac{1}{2\pi\nu}\frac{eE_0}{m} + v_0, \\ K_m &= \frac{1}{2}m\left[v_0 + \frac{1}{\pi\nu}\frac{eE_0}{m}\right]^2. \end{aligned} \tag{4}$$

If we knew the distribution-in-speed of the ions at the instant $t = 0$, and they were in a vacuum (i.e., never collided with atoms), we could predict their future motions. But they are not in a vacuum, and we do not know their distribution-in-speed at $t = 0$. What use, then, can be made of the equations?

Little or nothing, I am afraid, of such definiteness as to permit of exact and verifiable predictions about high-frequency discharges! After all, it is of the essence of a discharge that the phenomena occur in a gas where molecules and ions make collisions with each other; inferences drawn from the assumption that ions never collide with molecules are not likely to be close to truth. Nevertheless we may make some deductions of general value.

Thus: the expressions for the amplitude of the vibration of the ion, the constant speed of the point about which it vibrates, and the maximum kinetic energy acquired by the ion, all of them decrease when m is increased. All of them either are inversely proportional to m, or else involve a term inversely proportional to m. Therefore all of them are much smaller for positive ions than for free electrons. If a gas contains ions of both these kinds, the free electrons have by far the most energy, oscillate the farthest and drift the fastest. On this account we may often pretend that all the ions in the gas are stationary, excepting only the free electrons; and I shall often make this pretence.

Again: the fact that the center of oscillation of each ion drifts (except in a special case, presumably very rare) suggests that even in a vacuum it is impossible to confine the ions to any restricted part of the region between the electrodes.

Finally: the value given in (4) for the maximum kinetic energy of the ions, with the particular choice of zero for the value of v_0, may be taken as an indication of the order of magnitude of the energy which an ion might acquire in a gas not so dense but that occasionally it might run without a collision for a time as long as the duration of one oscillation. But for more than the order-of-magnitude we should certainly not rely on it!

Since allowance should be made for the collisions of the ions with molecules of the gas, let us do what is customary in physics: rush to the extreme, exaggerate the effect instead of neglecting it, and suppose that collisions are so frequent that their net influence upon a wandering ion amounts to a force of the type called "viscous" or "frictional"—a force proportional and oppositely directed to the velocity of the particle. Into the equation of motion we introduce a new term, so that it takes the form:

$$m(d^2x/dt^2) + g(dx/dt) = eE = eE_0 \sin nt \tag{5}$$

g standing for the new "coefficient of friction," and n being written for $2\pi\nu$ so as to avoid constant repetition of the factor 2π. One integration is sufficient to bring out the important point; we obtain:

$$\frac{dx}{dt} = \frac{eg}{m^2n^2 + g^2} E - \frac{me}{m^2n^2 + g^2} \frac{dE}{dt}. \tag{6}$$

If now we multiply each member of the equation by Ne — N standing for the number of ions per unit volume—each becomes equal to the current-density borne by ions in the gas.

Notice now an important thing about this expression for current-density. It consists of two terms, one of which is proportional to the fieldstrength E, while the other is proportional to the time-derivative of the fieldstrength—two terms, therefore, of which one is in phase with E and the other in quadrature with E. The first we may set down in the following equation:

$$\textit{Current-density in phase with fieldstrength} = \frac{Ne^2g}{m^2n^2 + g^2} E = \sigma E \tag{7}$$

and we may apply the name "conductivity" to the coefficient σ. The second must be treated differently. By itself it is not the whole of the current-density in quadrature with the fieldstrength; there is the displacement-current also to be taken into account, equal to $(1/4\pi)(dE/dt)$. Then we must write:

Current-density in quadrature with fieldstrength

$$= \frac{1}{4\pi}\left(1 - \frac{4\pi Ne^2m}{m^2n^2 + g^2}\right)\frac{dE}{dt} = \frac{\epsilon}{4\pi}\frac{dE}{dt} \tag{8}$$

and we may apply the name "dielectric constant" to the coefficient ϵ.

Say now that we can ionize a gas abundantly, as for instance by maintaining an intense direct-current discharge across it between auxiliary electrodes; and that we can apply a small oscillating voltage transversely, between a pair of electrodes facing one another like the plates of a condenser; and that we can measure, in the circuit containing this oscillating voltage and these condenser-plates, the components of current in phase and in quadrature with the field—then by equations (7) and (8), we can determine N and g. Or if we can measure either component by itself, and N in some independent way, we can determine g and test the equations by comparing its value with one derived from our knowledge of the flow of ions through gases in a steady field. I will describe how this is done, after giving equations for a more special and a more general case.

If the coefficient g is very large compared to mn—we shall later see that this is the same as saying: if an ion makes very many collisions with molecules during a single cycle of the applied field—the component of convection-current in quadrature with fieldstrength almost vanishes. The ions describe simple-harmonic vibrations about fixed centres (there is no possibility of a steady drift of the centre of vibration), the velocity being in phase with the field; the amplitude of vibration is easily shown to be given by this expression:

$$\text{Amplitude} = eE_0/2\pi\nu g = \mu E_0/2\pi\nu. \tag{9}$$

I take this opportunity of mentioning that the quantity e/g, the quotient by E of the drift-speed attained by the ion under a *steady* field of strength E, is called the "mobility" of the ion at the pressure in question, and is denoted by μ.

If the equation of motion of the ion includes in addition to the two terms in the left-hand member of equation (5) a third term fx (I have written it out with this additional term farther along in the article, as equation 26), the expressions for the current-components in phase and in quadrature with field strength involve the coefficient f. The formulae for σ and ϵ become the following:

$$\sigma = \frac{Ne^2g}{[(f/n) - mn]^2 + g^2} \qquad \epsilon = 1 - \frac{4\pi Ne^2[m - (f/n^2)]}{[(f/n) - mn]^2 + g^2}. \tag{10}$$

The term in question and its coefficient f would naturally be introduced if we were dealing with bound electrons, but it seems odd to postulate such a term in speaking of freely-moving electrons or ions; nevertheless there is reason to do so, as I will later mention.[2]

When a gas containing free electrons is traversed by electromagnetic waves, there is at every point an oscillating electric field having the frequency of the waves, and the electrons vibrate according to the foregoing laws. The waves are then absorbed—that is to say, their amplitude falls off exponentially as they progress—and their phase-velocity is altered. In the particular case in which g and therefore σ are negligibly small, there is no absorption, and the phase-velocity is altered in the ratio $1 : \sqrt{\epsilon}$—a fact which is described by saying that the index of refraction of the electron-populated gas is equal to $\sqrt{\epsilon}$. Since ϵ is less than unity according to equation (8), the waves go faster than they would if there were no free electrons roaming the gas. Formulae for index of refraction and absorption-coefficient, in the general case in which g and σ do not vanish, are given further along in this article (equations 15, 16). The transmission of radio waves through the atmosphere of the earth, in which free electrons are present in a concentration varying greatly with altitude, is much affected by this refraction.

---

[2] There is also an allowance for polarization which should be made if extreme accuracy is desired. See Appleton & Chapman, *l.c. infra.*

Experiments on Oscillation of Electrons in Ionized Gases

There is a method which, in principle, is adequate for measuring the dielectric constant and the conductivity of an ionized gas subjected to high-frequency vibrations; adequate therefore, in principle, for testing the expressions (7) and (8), for evaluating the quantities g and N which figure in those expressions. I will describe the method, first in the form in which it was applied to ionized air by Szekely, with results which are of some interest in themselves but mainly serve to illustrate how great a gulf intervenes between "adequate in principle" and "adequate in practice."

The gas—air at some pressure of the order of a few hundredths of a millimeter of mercury—was contained in a long tube, having electrodes near its ends whereby a direct current could be passed through the tube, maintaining the air in a state of intense ionization. Inside the tube and near its middle were a pair of parallel plane electrodes, rectangular in shape (18 by 42 mm.) and 4.5 mm. apart; to these the high-frequency voltage was applied; I will speak of them hereafter as "the ionization-condenser." When the direct current was flowing, a self-sustaining glow-discharge existed in the tube, and the ionization-condenser was submerged in the negative glow thereof.

The ionization-condenser was shunted by an adjustable condenser outside of the tube, and the two of them by an inductance and an adjustable resistance; all this constituted a circuit—I will call it hereafter the "high-frequency circuit"—coupled at one place to a source of high-frequency E.M.F., at another place to a detector. With such an arrangement, when one varies the capacity of the adjustable condenser and leaves everything else unchanged, the current in the circuit (measured by the detector) passes through a maximum at a certain setting of the capacity. At the attainment of this maximum, the system is said to be in resonance.

In Szekely's experiments, the adjustable condenser was set to produce resonance in three different conditions: in case (*a*) the ionization-condenser was entirely disconnected; in case (*b*) it was connected, but there was no direct-current discharge flowing in the tube, and the air was not ionized; in case (*c*) it was connected, and the air was ionized by the direct-current discharge. The values C_a, C_b, C_c of the adjustable capacity were measured at resonance under the three conditions; their differences give the values of the capacity of the ionization-condenser (plus that of its leads, which apparently is a large part of the total) when the gas is ionized and when it is not. Out of these data one may calculate the dielectric constant of the ionized gas. As for the conductivity: the values i_{mb} and i_{mc} of the high-frequency current at resonance are different in cases (*b*) and (*c*), smaller in the latter. After measuring i_{mc}, one may return to condition (*b*)—shutting off the direct-current discharge—restore the resonance, and reduce i_{mb} to the value just found for i_{mc} by adding resistance to the circuit. From the amount ρ of added resistance which is necessary to achieve this, one may calculate the resistance between the condenser-plates when the air is ionized; and thence—taking account of the size and shape of the plates and the distance between them—the conductivity of the ion-populated air.

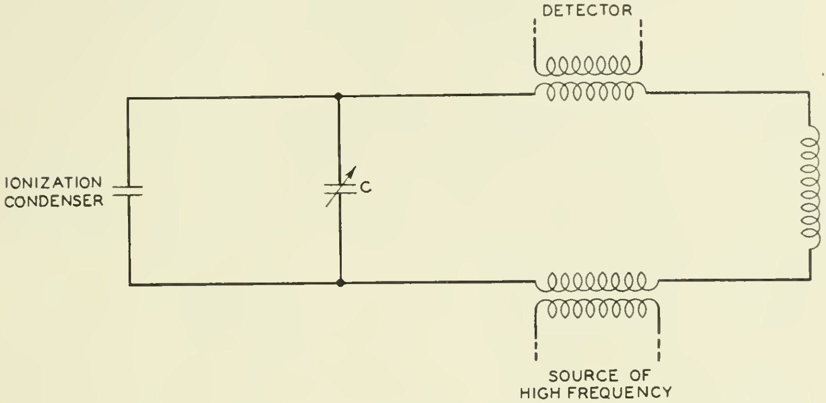


Fig. 2—Illustrating a method for measuring dielectric constant and conductivity of an ionized gas (between the plates of the "ionization-condenser").

Everything thus is apparently provided for testing the expressions and determining the constants derived from the theory of electrons oscillating in the ionized gas. But, the results of the experiments were not agreeable. True, in one important respect there was concordance with theory. Szekely measured the resistance of the ionization-condenser (or, to express it better, the resistance of the ionized air between the plates) for various values of the direct current sustaining the ionization, and various frequencies ranging from one to five millions. Plotting against the square of the frequency those values of resistance which belonged to one and the same strength of direct current, she obtained ascending straight lines (Fig. 3). Now, this agrees with the equation (7) supplied by the theory, according to which the reciprocal of the conductivity of an ionized gas is a linear ascending function of frequency squared, if N be constant;

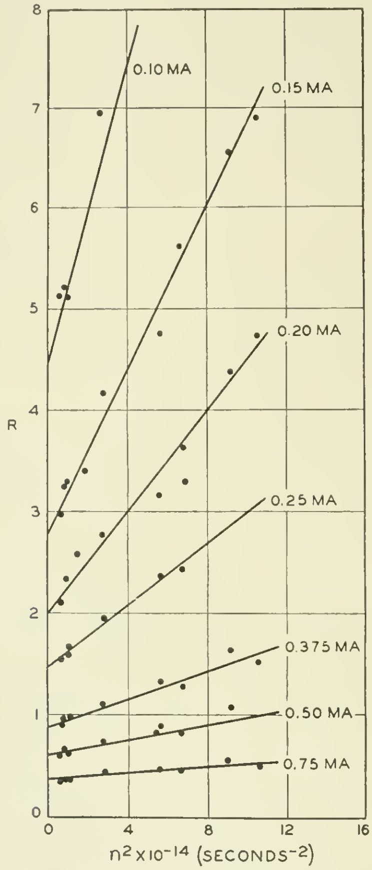


Fig. 3—High-frequency resistance of ionized rarefied air, plotted against square of frequency. (A. Szekely; *Annalen der Physik.*)

but the values computed from these curves for N and g are not very plausible, as I will later stress. What is much more serious: the capacity of the ionization-condenser is increased by ionizing the air; the dielectric constant of the ionized air is greater than unity, instead of less! The method therefore, adequate as it seems in principle, suffers from some defect or defects, which it is important to discover.

One of these defects was recognized towards 1925 by Appleton, who in previous work by the same sort of method had obtained the same bothersome result: when the air was ionized, the capacity of the ionization-condenser usually went up instead of down (although he, and van der Pol before him, did have the satisfaction of observing a decline of the capacity, when the strength of the direct current and hence the degree of ionization were relatively low). This however he explained by taking into account the space-charge sheaths of positive ions which form upon plates immersed in a strongly-ionized gas, or for that matter upon the walls of the tube containing the gas, if they are allowed to assume the potential which they naturally seek.[3]

If each of the two plates of the ionization-condenser is overspread by a positive-ion sheath of thickness x (x being less than half the distance d between the two plates, otherwise the sheaths would overlap) the system behaves not like a single condenser but like two in series. The capacity of each of the two, according to Appleton and Childs, varies inversely as x; their formula for each is $AK/3\pi x$, K standing for what they define as "the effective dielectric constant of the sheath" and A for the area of the plate; it differs from the customary formula for the capacity of a plane condenser—$AK/4\pi x$—because of the distribution of charge in the volume between the plate and the outer edge of the sheath. The capacity of the two in series is $AK/6\pi x$. If x is sufficiently small, this will be considerably larger than the capacity $A/4\pi d$ of the ionization-condenser as it was when the gas was not yet ionized; and the change in capacity occurring when ionization is started will consist chiefly of the substitution of the value $AK/6\pi x$, for the value $A/4\pi d$—even the influence of the layer of ionized gas between the outer surfaces of the sheaths will be minor.

If these ideas are correct, then when x is small the change of capacity of the ionization-condenser should vary inversely as x, provided we can vary x without altering K. Now when the potential V of the plates relative to the bulk of the ionized gas is varied, and the intensity of the ionization in the gas remains the same, the thickness of the sheath varies as $V^{3/4}$ while the current-strength across it remains unchanged.[4] If K remains the same, then the increment of capacity due to the ionization should vary as $V^{3/4}$; its logarithm should be a linear function of the logarithm of V, the line having slope $-3/4$. This inference was tested by Appleton and Childs, with the favorable result displayed in Fig. 4.

---

[3] "Electrical Phenomena in Gases," pp. 355–371.

The space-charge sheaths thus furnish an explanation for the unexpected sign of the change in capacity occurring when the gas is suddenly filled with ions—an explanation releasing us from having to assume that the dielectric constant of the gas rises above unity, in contradiction to the simple theory.

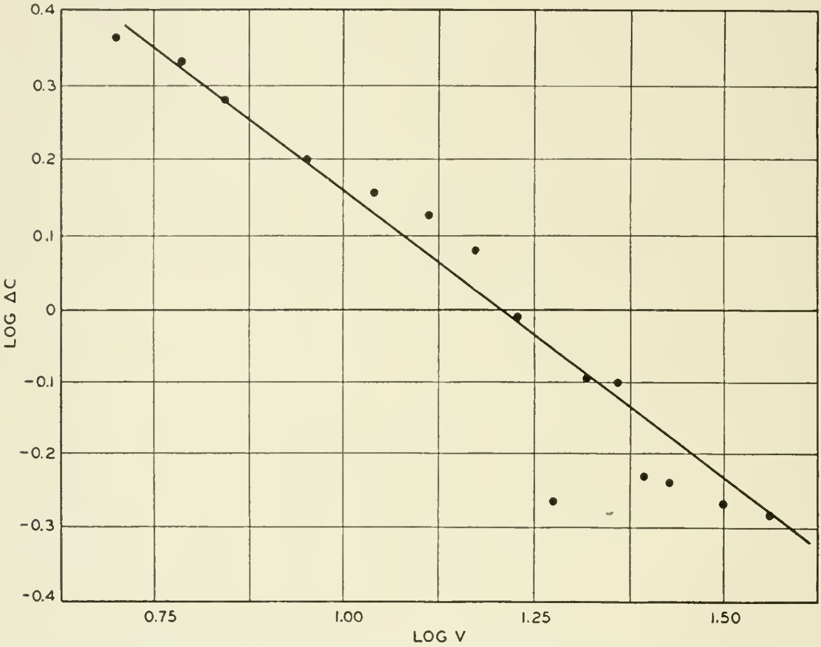


Fig. 4—Alteration in capacity of a stratum of ionized gas, ascribed to the formation of positive-ion space-charge-sheaths along the surfaces of the electrodes bounding the stratum. The slope of the line is $-\frac{3}{4}$. (E. V. Appleton & E. C. Childs; *Philosophical Magazine.*)

Could the sheaths be eliminated? If x is allowed to increase indefinitely, several things happen; in particular, the assumption that the thickness varies as $V^{3/4}$ and the assumption that $K/6\pi x$ is much larger than $1/4\pi d$ depart further and further from the truth, and when x passes $d/2$ the sheaths overlap and the conditions change entirely. Other things being equal, x increases as the number of ions per unit volume of the gas outside the sheath decreases; and it is actually found that as the direct current maintaining the ionization is reduced, the increase of capacity due to the ionization is also reduced, passes through zero, and becomes a diminution as the simple theory indicates.

---

[4] "Electrical Phenomena in Gases," equation (182), page 360. The equation is the familiar "three-halves-power equation," which derives its usual name from the relation between voltage and current-density with which we are not here concerned. It is valid only if the ions mostly cross the sheath without collisions, an important restriction.

One might on the other hand get rid of the sheaths altogether, by adjusting the (mean) potential of the plates to so low a value that x vanishes, instead of leaving it to seek its own level. This has recently been done by Childs, though for some reason he chose to study not the dielectric constant, but the conductivity of the ionized air. A galvanometer was coupled into the high-frequency circuit, which was tuned to resonance when the mean potential of the plates had been so adjusted that the visible sheaths had just vanished. The reading of the galvanometer was taken; the current maintaining the ionization (it was 300-cycle A.C., instead of D.C.) was discontinued, and various high resistances were connected in parallel with the condenser plates, until the galvanometer resumed its former reading; the value of resistance then existing was taken as that of the ionized air between the condenser-plates. The edge-correction of the condenser was determined by filling the tube with alcohol of known conductivity; it was found that the conductance of the condenser had to be multiplied by 0.57 to be converted into conductivity of the medium between the plates. The values of conductivity (I will presently state the conditions more precisely) were of the order of 10^{-14} E.M.U.

A value of conductivity by itself, obtained at a single frequency, is theoretically the value of a combination of N and g; to determine either, without the aid of a simultaneous measurement of dielectric constant, one must have an independent value of the other. Childs evaluated N by the Langmuir probe-method.[5] Working with air at 1 mm. pressure and a frequency about one million, for three values of the 300-cycle A.C. maintaining the ionization (5 ma., 10 ma., 15 ma.) he obtained three values of the electron-concentration N (they were 9.6, 24 and 36 times 10^7 per cc.), and substituted them, not into the general equation (7) for σ, but into an approximate form thereof:

$$\sigma = Ne^2/g, \tag{11}$$

believing g to be so large that mn is small by comparison. The values of g thus obtained were 2.5, 3.1 and 2.5 times 10^{-18}. Despite

[5] "Electrical Phenomena in Gases," pp. 351–352.

the large differences between the values of N, those of g show no trend, which is as it should be.

Another way of determining g without bothering about N has just been carried into effect by Appleton and Chapman. Returning to equation (7): one readily sees that, if it be possible to vary g without altering N, the conductivity should pass through a maximum when $g = mn$. Now according to the simple interpretation the coefficient g, being the measure of a sort of friction which the electrons suffer when travelling through a gas, should increase with the pressure. If therefore one can vary the pressure of the gas sufficiently while applying the same high frequency, and maintaining the same degree of ionization, one should find the conductivity passing through a maximum. Moreover, anything proportional to the conductivity of the gas should pass through a maximum, anything inversely proportional to the conductivity should pass through a minimum. This conveniently makes it unnecessary to measure the conductivity (or anything else) absolutely; it is sufficient to choose something which varies (say) inversely with conductivity, plot it as function of pressure, and locate the minimum of the curve. For the pressure corresponding to this minimum, the value of g is equal to the product of the mass of the electron by the applied frequency.

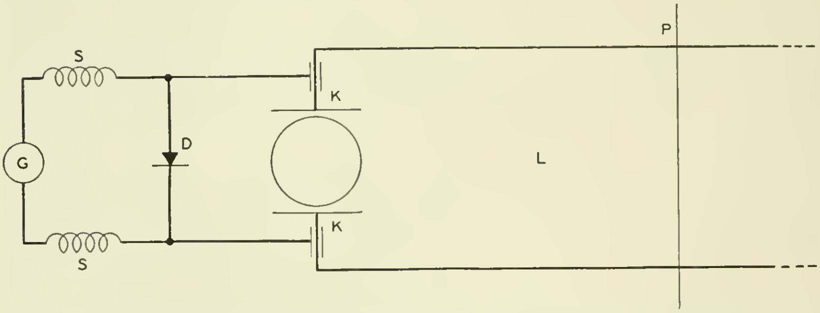


Fig. 5—Illustrating another scheme for estimating the dielectric constant and conductivity of an ionized gas (in the tube between the plates K), with Lecher wires and movable bridge P replacing the adjustable condenser of Fig. 1. (E. V. Appleton & F. W. Chapman; *Proc. Phys. Soc. London.*)

The apparatus and the circuit of Appleton and Chapman are shown in Fig. 5; as one notices, the plates of the ionization-condenser are now outside the tube and the ionized gas, and the high-frequency circuit now comprises long parallel wires, known (after the man who first used them as portions of high-frequency circuits) as Lecher wires. The current through the galvanometer to the left (shunted by a galena detector) is the measured quantity; the change which occurs in the reading of the galvanometer, when the discharge and the ionization commence in the tube (resonance being restored in the circuit, by shifting the "bridge" across the wires), is the quantity which supposedly varies inversely as conductivity. In the curves of Fig. 6, one sees the minima. Such curves were plotted for four frequencies—340, 500, 550 and 625 millions—and the maxima occurred at the pressures 0.08, 0.11, 0.12 and 0.15 mm. Hg. The values of mn corresponding were 1.9, 2.8, 3.1 and 3.5 times 10^{-18}; these by the theory are the values of g for the corresponding pressures. They are roughly proportional to the pressure, as they should be. (Incidentally, the resonance-frequency of the system was shifted, when the ionization commenced, in the proper sense—the sense corresponding to a dropping of the dielectric constant below unity.)

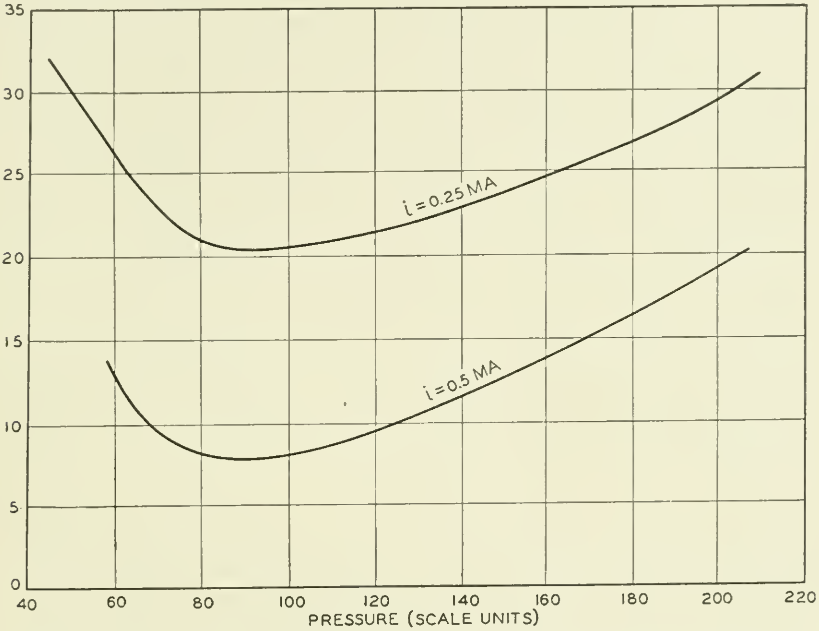


Fig. 6—Evidence of a maximum of conductivity occurring when the frequency is related in a particular way to the frictional coefficient g. (Appleton & Chapman.)

Before considering further data, let us compare these values of g with those deduced from observation (or theory) of electrons drifting through the same or a similar gas under a constant field. In such a condition, they attain a "terminal drift-speed" u given by the equation,

$$gu = eF, \tag{12}$$

F standing for the fieldstrength. Despite the aspect of the equation, u is not proportional to F; the coefficient g must be regarded as a function of F (over wide ranges of fieldstrength it is proportional to the square root thereof). Though this is a fact of experience, it will be helpful to develop the theory to some extent.

Picture the drifting of electrons through a gas in the customary (though far too primitive) way. Imagine the gas as a congregation of elastic spheres, with which the electrons make elastic impacts. When these latter enter the region where the constant and uniform field is applied, they are speeded up in the direction of the field; but owing to the deflections and the losses of energy which they suffer at their impacts, their velocities become and remain almost isotropic, and their average energy approaches but does not surpass a certain limiting value determined by the fieldstrength. After they have progressed sufficiently far, they form an "electron-gas" mingled with the atoms of the material gas; this electron-gas drifts slowly along towards the positive electrode, but its individual corpuscles have (as a rule) random velocities tremendously in excess of that comparatively modest drift-speed, just as the molecules of the air have random velocities many times greater than the speed of the wind. Let me denote the drift-speed by u, the mean speed of the random motions of the electrons by ω. Now it may easily be shown [6] that if the simple picture of the molecules as big elastic spheres and the electrons as little ones is acceptable, u and ω and the mean-free-path of the electrons l are related by the equation:

$$u = \theta eEl/m\omega, \tag{13}$$

θ here standing for a numerical factor not very different from unity, the exact value of which depends on the underlying assumptions made in the statistical analysis of the motions of molecules and electrons. The value 0.8 for θ is probably as good as any, but it would be pointless to spend time deciding between different values, since the molecules cannot be represented exactly as elastic spheres, and the degree of their deviation from this simple model would affect the quantity θ.

Now comparing the last two equations, we find:

$$\begin{aligned} g &= m\omega/\theta l \\ &= m/\theta\tau = mZ/\theta, \end{aligned} \tag{14}$$

τ and Z here representing respectively the mean duration of the free flight of an electron between consecutive impacts, and the number of impacts made by an electron in unit time. (I introduce these symbols because they are used in some of the original articles, not because they add anything to the fundamental ideas.)

---

[6] "Electrical Phenomena in Gases," pp. 174–175.

The value of ω may be determined by the Langmuir probe-method; values for l, the mean-free-path of the electrons, are supplied by various methods of varying reliability. Childs, in his experiment which I have quoted, obtained with the probe the three values 1.3, 1.5, 1.7 times 10^8 (cm./sec.) for ω, these corresponding to the three cited values of the current maintaining the ionization; combining these with estimates [7] of l, he obtained for $m\omega/l$ the values 3.2, 3.8, $4.3 \cdot 10^{-18}$. Comparing these with the values found for g, and remembering the manifold chances of faults in the assumptions, one is favorably impressed with the agreement.

Could we not evaluate g directly, by measuring the drift-speed u of the electrons exposed to a certain constant fieldstrength F, and forming the ratio of eF to u to which, according to equation (12), g is equal? Here we must be careful. According to equation (13), g depends on the vivacity of the random motions of the molecules, whereof ω is the mean speed; now, ω depends on the fieldstrength; we must select such a value of F, that the random agitation of the electrons shall be the same as it is in the actual gas on which the high-frequency field is imposed. Now this actual gas is subjected to a constant field, that which maintains the ionization. It is natural and simple to assume, that this field controls the value of ω, the effect of the high-frequency field on the mean speed of random agitation of the electrons being presumably slight. The measurement of the drift-speed should therefore be made in the very gas under the very same constant field, in which the high-frequency phenomena are observed.

Let me denote by i the steady current along the length of the tube, due to the constant field; by N, the number of electrons per unit volume; by u, their drift-speed; by A, the cross-sectional area of the tube. It can then be readily seen that i is equal to $NeuA$.[8] The method is consequently simple in principle; in practice, the chief difficulty apparently is that N is not the same near the walls of the tube as along its axis, so that probe-measurements should be made at a number of distances from the axis and the results averaged. Appleton and Chapman determined N at a single point of the tube; the value of the ratio eF/u came out equal to $2.3 \cdot 10^{-18}$—again, a remarkable agreement with the values deduced from the high-frequency phenomena![9]

---

[7] Apparently he multiplied by $4\sqrt{2}$ the value of the mean-free-path of molecules of nitrogen, this being the simple gas presumably most like air; at any rate he used the value .036 cm.

[8] *Electrical Phenomena in Gases*, p. 207, pp. 232–233; on the two latter pages I describe Killian's application of the method to mercury vapor.

(It is rather surprising to realize that in such discharges, we must not conceive the electrons as moving with a steady velocity along the axis of the tube, on which a sinusoidal oscillation parallel to the field is superposed. If we could follow the wanderings of an individual electron, the sinusoidal oscillation would be as little obvious as the steady drift. Only the rapid zigzag motions of the corpuscles would be conspicuous; what I have been calling a vibratory motion is, in truth, only a slight bias of these random flights, just as the apparent steady drift is itself a slight bias.)

Another experiment, capable in principle of testing the expressions for dielectric constant and conductivity and of giving the values of N and g, consists in sending electromagnetic waves of the frequency desired across a stratum of ionized gas, and measuring their index of refraction and their absorption in the stratum. The former of these two has not, so far as I know, been measured; but apparatus for determining the second (it is that of Hasselbeck's experiment) is shown in Fig. 7: one sees the paraffin lenses which convert a diverging beam of Hertzian waves into a parallel beam which is sent through the ionized gas, and others which reconvert the parallel beam into one which converges upon a bolometer. Part of the beam is reflected from a semi-transparent mirror onto another bolometer, so that the ratio of the intensities of the waves before and after the passage through the gas can be determined without regard to fluctuations.

The formulae for refractive index n_0 (I add the subscript because of having already used n, the conventional symbol, for another purpose) and absorption-coefficient k are familiar to everyone who has studied the theory of absorption and reflection of light by metals, for in the classical theory of metals such a substance is conceived exactly as we are now conceiving an electron-populated gas. We have:[10]

$$n_0^2 - k^2 = \epsilon, \qquad n_0 k = 2\pi\sigma/n, \tag{15}$$

and putting for ϵ and σ the expressions (7, 8), we get:

$$n_0^2 - k^2 = 1 - \frac{4\pi Ne^2 m}{m^2n^2 + g^2}, \qquad n_0\, k = \frac{2\pi}{n}\,\frac{Ne^2 g}{m^2n^2 + g^2}, \tag{16}$$

solving which equations for k, and putting $m/\theta\tau$ for g (according to equation 14) and n_1^2 for the combination $4\pi Ne^2/m$ (we shall meet it later in another meaning), and α for the ratio n_1/n, we obtain:

---

[9] Similar observations by Jonescu and Mihul on air and on hydrogen, subjected in some cases to magnetic field, have recently been published, but with scant detail.

[10] See for instance P. Drude, "Treatise on Optics" (page 361 of the English translation).

$$k^2 = \frac{1}{2(1 + n^2\theta^2\tau^2)} \{\sqrt{[1 + n^2\theta^2\tau^2(1 - \alpha^2)]^2 + n^2\theta^2\tau^2\alpha^4} + [1 + n^2\theta^2\tau^2(1 - \alpha^2)]\}. \quad (17)$$

By measuring k with waves of a single frequency, and measuring in addition either of the two quantities N and g, it is possible to evaluate the other of the two by means of this equation; provided the equation is correct, a presumption which should be tested by making the measurements at several different frequencies.

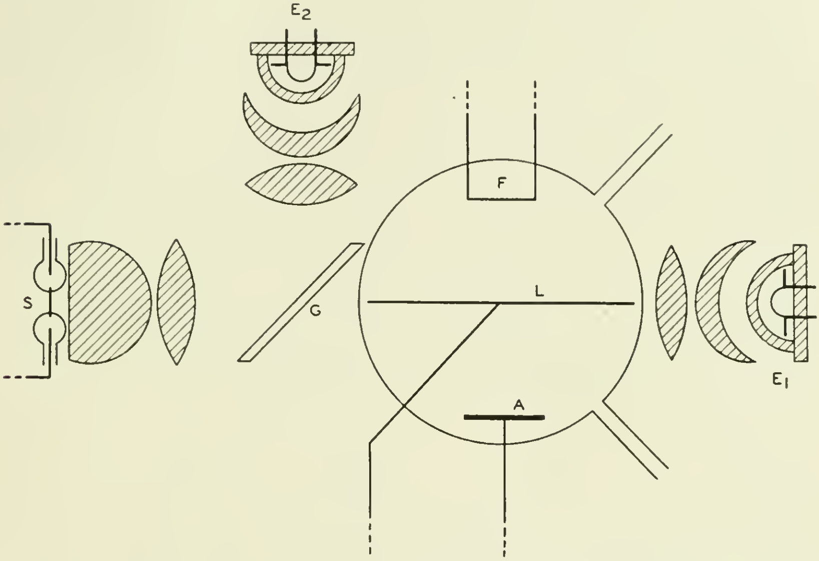


Fig. 7—Apparatus for measuring the absorption of Hertzian waves by ionized gas. S, source of waves; G, semitransparent mirror; E_1, E_2, receivers; A, F, anode and cathode of ionizing discharge; L, wire probe. (W. Hasselbeck; *Annalen der Physik.*)

Dänzer sent 4-cm. waves (of frequency $7.5 \cdot 10^9$)through a stratum of neon-helium mixture excited by a low-frequency discharge; measured the absorption, and measured also g by the method later employed by Appleton and Chapman. For N he then computed the value $1.3 \cdot 10^{12}$ (electrons per cc.)—a value of which, in the lack of further knowledge concerning electrons in this gas (he does not even state the pressure) one can only say that it is probably of the right order of magnitude. Much more extensive was the work of Hasselbeck, who used various frequencies ranging from 4.8 to 1.44 times 10^9, and introduced a probe into the gas in order to measure N and the mean speed ω of the electrons by Langmuir's method—it was a wire stretching clear across the globular tube, being made of this form so that it might give average values appropriate to the whole of the discharge. The best of his data refer to neon-helium mixture at various pressures of the order of a few tenths of a millimeter; there are data also for argon and nitrogen.

From equation (16), k should be proportional to N so long as n_0 does not depart too much from 1 and other things remain unchanged. Hasselbeck varied N by varying the current maintaining the ionization, measured it with the probe and measured the absorption. So long as the pressure of the gas was below a certain value, of slightly over 0.1 mm. Hg, the curves of k vs N were indeed straight lines rising from the origin (Fig. 8), though at higher pressures they were concave-upward. As for the curves of absorption *vs* frequency, some display maxima in the range over which they are plotted, some do not; but it is possible to suppose that each is characterized by a single maximum, which moves toward higher frequencies when the pressure or the ionization is augmented, and in some of Hasselbeck's experiments happened to be out of the range.

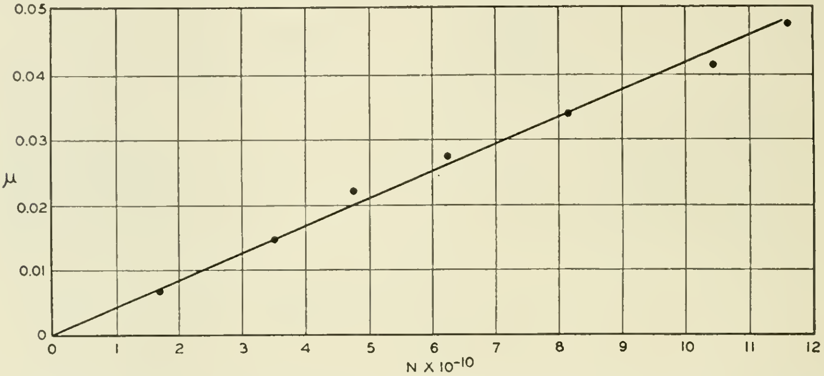


Fig. 8—Showing the proportionality of absorption of Hertzian waves to number of free electrons per cc. of absorbing gas (neon-helium mixture, pressure 0.18 mm. Hg). (W. Hasselbeck; *Annalen der Physik.*)

The comparison of the absorption-*vs.*-frequency curves with so complicated an expression as that of equation (17) is no simple matter, nor is Hasselbeck's description of his procedure entirely clear. It appears, however, that the curves obtained in neon-helium mixture at a pressure of 0.87 mm. Hg depart but little from the form of the expression on the right of (17), and that what departure there is may not be significant. However the curves obtained at the much higher pressure of 21 mm. Hg differ very seriously from the theoretical form.

Reverting to the question of testing equation (8) for the dielectric constant, that is, for the current in quadrature with the voltage: one might try the test with a gas-free space, populated by electrons derived from a hot filament or from some other source. This in fact has been tried, the electrons being shot across the interspace between the plates of the "ionization-condenser," in directions nearly parallel to the plane of these plates; from the (controllable) speed of these electrons, and from the charge which they bear per unit time to a collector located beyond the condenser, one may estimate the number-per-unit-volume heretofore denoted by N.

Since the density of the gas may easily be made so low that few electrons collide with even one atom between the condenser-plates, one expects the coefficient g to vanish. Strictly speaking, it does; and yet there is in effect a component of current in phase with voltage. The electrons, being pulled aside by the transverse high-frequency field as they traverse the condenser, acquire kinetic energy; and this absorption of energy from the high-frequency circuit produces the same reaction, and is measured in the same way, as an ordinary conductance. Benner, who developed these ideas, derived (with certain simplifying assumptions) these expressions for the coefficients ϵ and σ of a cloud of streaming electrons which individually take the time T to cross the condenser:

$$\epsilon = 1 - 4\pi\frac{Ne^2}{mn^2}\left(1 - \frac{\sin nT}{nT}\right), \qquad \sigma = \frac{Ne^2}{mn^2T}(1 - \cos nT), \quad (18)$$

and they have been tested with satisfactory results by Jonescu and Mihul.

We turn now to the possibility for which preparation was made in equations (10); that an electron in an ionized gas may experience, in addition to the force due to the applied field (eE) and the quasi frictional force due to the gas ($g\dot{x}$) yet a third force proportional to its distance from some fixed point. One is accustomed to postulate this last for electrons bound to molecules or atoms. (Such electrons, by the way, contribute to the current-component in quadrature with field strength, so that the dielectric constant of a gas in absence of ionization is not quite unity as I have been writing). But to imagine such a force acting on free electrons must seem strange—as if one were denying them the quality of freedom. Still, for a cloud of electrons mingled with positive ions there is such a force, and therefore a "natural frequency"; and another natural frequency in addition, if there happens to be a magnetic field.

Considering the former first: suppose that initially we have a uniform distribution of electrons, N per unit volume; and that at a certain moment it is suddenly distorted, by shifting every particle a distance ξ parallel to the x-direction, this distance being a function of the original position x of the particle. Fix attention on a column of unit cross-section; initially, between two planes x and $x + dx$, there were Ndx electrons; they suddenly move over and occupy the space between the planes $x + \xi$ and $x + dx + \xi + (d\xi/dx)dx$, so that the density between these two planes, or let me say the number of electrons per unit volume at $x + \xi$, is given by the formula:

$$N'_{x+\xi} = N\left(1 - \frac{d\xi}{dx}\right). \tag{19}$$

We now introduce Poisson's relation between net density ρ of electric charge and space-derivatives of field strength. As by assumption all shifts of electrons are parallel to the x-direction, so also is the field-vector; we put X for its magnitude, and write Poisson's equation thus:

$$dX/dx = 4\pi\rho. \tag{20}$$

If the electrons were the only charged particles, we should have to put $N'e$ for ρ; and there would be a field of strength different from zero and varying from place to place, even when the distribution of the electrons was uniform. If however there is also a uniform distribution of positive ions, N per unit volume, and this is not affected when that of the electrons becomes non-uniform, then for ρ we need set only the second term on the right-hand side of (19), multiplied by e; we get

$$dX/dx = -4\pi Ne(d\xi/dx), \tag{21}$$

and integrating, with the boundary-condition [11] $X = 0$ at $\xi = 0$,

$$X = -4\pi Ne\xi, \tag{22}$$

so that an electron shifted from its original location, by virtue of such a mass-distortion of the formerly uniform distribution, is indeed subjected to a restoring force proportionate to its shift.

Putting $m(d^2\xi/dt^2)/e$ for X in equation (22), we get:

$$d^2\xi/dt^2 = -(4\pi Ne^2/m)\xi = -n_1^2\xi, \tag{23}$$

showing that there is a tendency to oscillations—"plasma-electron oscillations," as Tonks and Langmuir call them—of frequency ν_1 thus given:

$$\nu_1 = n_1/2\pi = \sqrt{Ne^2/\pi m} = 8980\sqrt{N}. \tag{24}$$

---

[11] This seems to be demanded by symmetry if ξ is a sinusoidal function of x: otherwise the case is more obscure.

This expression is strictly valid only when the gas does not interfere at all with the motion of the electrons; otherwise the frequency is reduced, in the same way as the natural frequency of a pendulum or a circuit is lowered by damping. The equation of motion of the electrons in a high-frequency field is as follows:

$$m(d^2x/dt^2) + g(dx/dt) + fx = eE \sin nt; \qquad f = 4\pi Ne^2. \quad (25)$$

The solution has already been indicated (equations 10).

Attempts to discover this natural frequency have been made in two ways: by examining curves of σ or ϵ or other correlated quantities plotted against frequency or against degree of ionization, and by searching for electromagnetic waves due to oscillations arising of themselves in highly-ionized gases.

The most thorough experiments by the former way are due to Tonks. He placed a tube containing a mercury arc between condenser-plates attached to long parallel wires, these being crossed by a movable bridge including a thermocouple, and coupled to an oscillating circuit. After establishing fixed values of the frequency and the current-strength in the latter circuit, he shifted the bridge until the thermocouple reported a maximum of current, and measured this maximum value; it and the shift of the bridge (the zero from which this latter is measured is unimportant) were plotted as functions of the current-strength in the mercury arc, which controls the number of free electrons per cc. A natural frequency is indicated by a minimum in the former of the curves, and in the latter curve a peculiar crinkle, similar to that which appears in a dispersion-curve in the neighborhood of an absorption-frequency, and in the curve with black dots in Fig. 9.

Embarrassingly it turned out that there were two, and indeed sometimes three, minima in the one curve and crinkles in the other. To these the corresponding values of N were correlated, being obtained by the Langmuir probe-method. The comparison with theory may then be made in either of two ways: by putting the value of the applied high frequency into equation (24), computing N, and comparing it with the observed values of N at the minima; or alternatively, by putting the observed values of N into equation (24), computing ν_1, and comparing its values with that of the applied frequency ν. As an example of the result of the first procedure: in one experiment, the applied frequency was $1.59 \cdot 10^8$; this should coincide with the plasma-electron frequency, resonance should occur, when $N = 0.63 \cdot 10^9$ electrons per cc.; the values of N at the two minima which were observed were 0.27 and 0.86 times 10^9. The latter is illustrated by a graph in Tonks' article; it turns out that when the comparison is made at various frequencies between $1.59 \cdot 10^8$ and $3.66 \cdot 10^8$, the curves of ν_1 *vs.* N agree only fairly with those of ν *vs.* N; where ν_1 varies as the square root of N, ν varies as the powers 0.42 and 0.45 for the two minima regularly appearing.

Another instance of a crinkle in a curve of the second type aforesaid —a curve in which the shift of the bridge across the Lecher wires, necessary to restore resonance in the circuit which includes the ionization-condenser, is plotted against N or something varying with N—has been published by Appleton and Chapman. It refers to ionized air; the applied frequency is $3.75 \cdot 10^8$, the value of N is $1.62 \cdot 10^9$, and these two stand to one another in substantially the relation of equation (24). Yet other instances have been published by H. Gutton, and these offer difficulties.

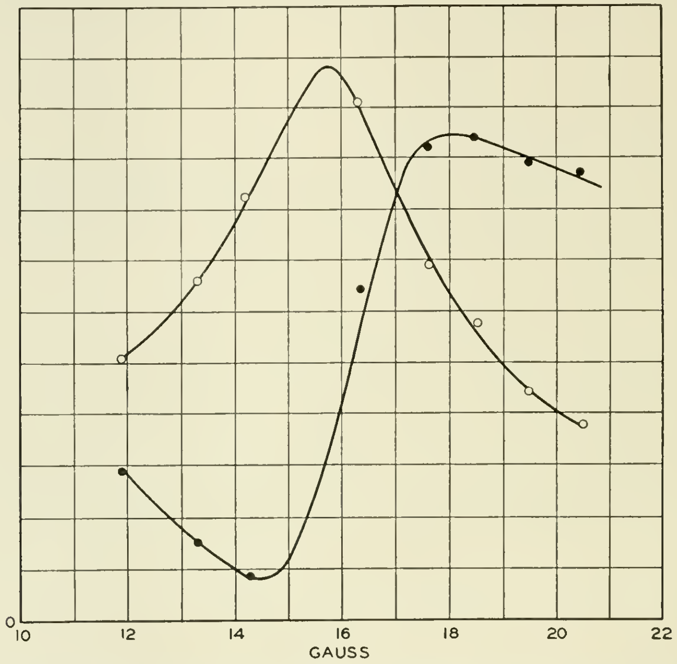


Fig. 9—Illustrating variation of dielectric constant (curve with full circles) and conductance (curve with hollow circles) of ionized gas in the neighborhood of a natural frequency, here due to a constant magnetic field. (S. Benner; *Naturwissenschaften.*)

Gutton in his later work (with ionized hydrogen at the low pressure of 0.0004 mm. Hg) followed the method which I have just described, measuring the shift y of the Lecher bridge required to restore resonance, and the mean-square current I^2 traversing the Lecher bridge at resonance. He did not measure N, but something probably (though not certainly) proportional to N, the factor of proportionality not known: the direct current i flowing across the gas between two probes inserted on opposite sides of the tube and maintained at a constant P.D. Plotting y and I^2 against i, he observed a crinkle in the former curve, a minimum in the latter—evidence of a natural frequency at a particular value of electron-concentration.

In his earlier work (which I mention because the curves are often reproduced) Gutton connected the plates of the ionization-condenser to one another through a thermocouple, and to this circuit coupled an oscillator of which the frequency could be varied. For each of a number of values of the ionizing-current in the discharge-tube he varied the frequency until resonance was declared by a maximum of the mean square of the current in the thermocouple; he measured this maximum I^2 and the wave-length λ of the oscillations. Not having any quantitative measure of the ionization against which to plot I^2 and λ, he plotted one against the other—I^2 as ordinate, λ as abscissa—and obtained curves of the curious appearance shown in Fig. 10, in which the arrow indicates the sense in which the ionization increases along each curve. The start is made from the wave-length (408 cm.) at which the system is in resonance when the gas is not ionized. At low pressure, the curve bends first to the left and downward; this signifies that as the ionization increases the dielectric constant of the gas is falling and the conductivity rising, as by the theory they should. Then at an unknown but seemingly sharply-marked value of ionization, the curve bends sharply to the right; and this signifies the same as the crinkle and the minimum in the other more-fully-comprehended curves which I have been discussing.

It is still dubious whether the natural frequency so revealed is that which the foregoing theory predicts. Neither in his earlier nor in his later experiments did Gutton measure N (the estimate of its value which he once makes is derived in an indirect and fallible way). Measuring the values of i at which the resonance appeared in his later work, and comparing them with the corresponding values of the frequency, he found ν^2 proportional to $i^{3/4}$; the range of frequencies was comparatively small (not quite 4 : 1) but if the result is certain and i is truly proportional to N, equation (24) is contradicted. Appleton and Chapman testify that they found the resonance-effect observed by Gutton, but at values of N (which they measured by the Langmuir method) entirely too great to permit of regarding it as the plasma-electron frequency, from which they believe it to be distinct.

As for the search for spontaneous oscillations probably having the proper plasma-electron frequency, it is at present even less advanced.

Electromagnetic waves of high frequency often emanate from gases intensely ionized by a flow of direct current, and their wave-lengths have several times been measured, but for one reason or another the comparison with equation (24) was not and cannot be made. In a mercury-vapor arc Tonks and Langmuir detected oscillations of frequencies nearly as high as 10^9, accompanied by others ranging downwards as low as 10^6; the former, they believe, were plasma-electron vibrations; but unluckily they were unable to estimate N with any degree of exactness.

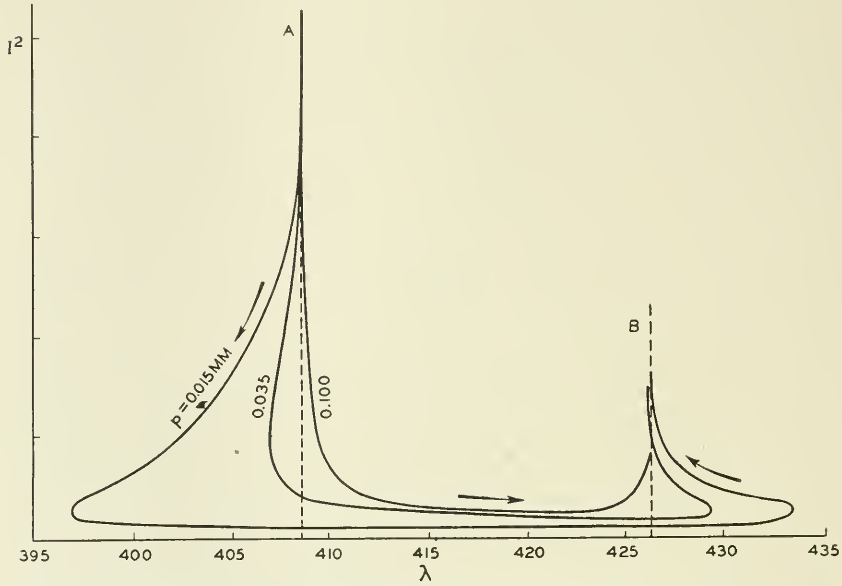


Fig. 10—Correlation of quantities serving as measures of high-frequency conductance and dielectric constant of rarefied ionized hydrogen. (H. Gutton; *Comptes Rendus.*)

If a region populated with free electrons is pervaded by a constant magnetic field, yet another natural frequency exists. Say, to begin with, that there is no electric field; then any moving electron, instead of continuing in a straight line, is constrained to describe a spiral path (the axis of the spiral being parallel to the magnetic field). The velocity of the electron affects the curvature of the helix, *but not the time of traversing a single winding* thereof. The reciprocal of this time, the number of windings described by any moving electron per second, is the "natural frequency" aforesaid, and is given by the formula:

$$\nu_H = n_H/2\pi = eH/2\pi mc, \qquad (26)$$

H standing as usual for the magnetic field strength. Since this frequency remains the same however much the speed of the electron may change, and therefore is the same for all the electrons in the region in question and for all values of electric field strength, we may expect it to be important when an ionized gas (or a volume containing free electrons but no atoms) is exposed to a high-frequency field; in curves of dielectric constant and conductance vs. frequency, we may look for peculiarities when $\nu = \nu_H$. The precise theory, I must add, is not simple when collisions of electrons with atoms must be taken into account (the fundamental equations were given long ago by Lorentz); but under certain restrictions—according to Appleton and Childs, the number of electrons per cc. and the number of collisions per cycle of the high frequency must be kept under certain limits—it leads to the inference that there should be a maximum of conductance at $\nu = \nu_H$.

This maximum is manifested by the sharp and striking minimum seen near the middle of the curve in Fig. 11. For obtaining these data, Appleton and Childs had an arrangement similar in the main to that of Fig. 5, excepting that the detecting galvanometer was coupled (through an amplifier) between the far ends of the wires of the bridge, and there was a magnetic field of adjustable strength parallel to the axis of the tube. Though I have spoken thus far of what should be observed when H is held constant and ν is varied through the value given by equation (26), it is more convenient in practice to hold the frequency constant and vary H through the corresponding value. The curve of Fig. 11 is accordingly a curve of galvanometer-reading

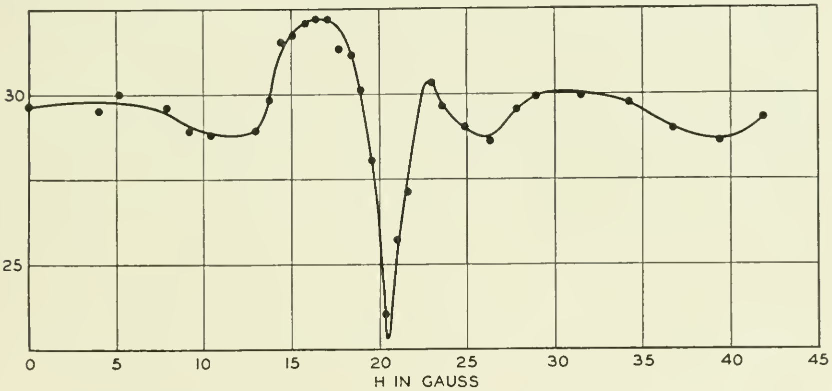


Fig. 11—Evidence of a natural frequency produced in an ionized gas by a constant magnetic field. (Appleton & Chapman.)

vs. H; the remarkable minimum occurs at a value of H departing by less than 2 per cent from $2\pi mc\nu/e$. These data were obtained at frequency $5.46 \cdot 10^8$; others observed at $2.96 \cdot 10^8$ and $3.45 \cdot 10^8$ yielded agreements almost as good, or better.

Maxima of conductivity of ionized air have lately been found by Jonescu and Mihul, at the predicted values of field strength H, for various frequencies of the order 10^8. Earlier Benner, in a note deplorably brief, showed not only a curve of conductivity (or rather, of something proportional to conductance) displaying a maximum, but a curve of dielectric constant displaying a crinkle like that of a dispersion-curve in the neighborhood of a region of anomalous dispersion. These are the curves of Fig. 9, already introduced into this article to illustrate how these quantities vary in the vicinity of a natural frequency. The "ionization-condenser" of Benner's experiment consisted of the grid and plate of a triode, the space between them populated with electrons emitted from the filament; it is to be inferred that the tube contained some gas, but unfortunately nothing definite is said about the kind or amount. The maximum of the one curve and the crinkle in the other occurred at a value of H some twelve per cent higher than the predicted value aforesaid.

On applying a longitudinal magnetic field of about 21 gauss to the tube of ionized hydrogen with which he had observed the peculiar natural frequency mentioned above, H. Gutton found this one replaced by two, well marked and well separated, one being shifted toward higher frequencies from the original value and one toward lower. The same phenomenon has been observed by Tonks, in his studies of the natural frequencies which he identifies with the plasma-electron oscillations predicted by the theory culminating in equation (24). The doubling of the resonance is analogous to the Zeeman effect, and is amenable to theory.

The earth's magnetic field imprints a natural frequency upon the electrons populating the air, in particular the upper strata thereof; this affects the transmission of radio waves in curious ways, which were foretold in this journal seven years ago by Nichols and Schelleng, and in England by Appleton and Barnett.

In the second part of this article, I will treat of the conditions under which a high-frequency field may initiate and maintain a luminous discharge in a gas, and of the laws of these discharges.

References

The theory underlying equations (1) to (4) is well known; it is given in considerable detail by G. Mierdel, *Ann. d. Phys.*, **85**, 612–640 (1928), with some illustrative curves, and in several other papers. The theory underlying equations (5) to (10) apparently originated, like so much else in electronics, with H. A. Lorentz; it was developed with especial reference to the transmission of radio waves by W. H. Eccles, *Proc. Roy. Soc.*, **A 87**, 79–99 (1912); later enunciations, with various modifications, may be found in the papers of J. Salpeter, *Phys. ZS.*, **14**, 201–203 (1913); E. V. Appleton & F. W. Chapman, *Proc. Phys. Soc. Lond.*, **44**, 246–254 (1932), and in several papers cited in the next paragraph.

Experiments designed to measure dielectric constant and conductivity of ionized gas for high-frequency fields have been performed by A. Szekely, *Ann. d. Phys.* (5), **3**, 112–132 (1929); E. V. Appleton & E. C. Childs, *Phil. Mag.* (7), **10**, 969–994 (1930); Appleton & Chapman, *l.c. supra;* E. C. Childs, *Phil. Mag.* (7), **13**, 873–887 (1932); H. Gutton, *Annales de physique* (10), **13**, 62–129 (1930); H. Gutton & J. Clement, *C. R.*, **184**, 441–443, 676–678 (1927); V. Jonescu & C. Mihul, *C. R.*, **192**, 343–345 (1931.)

Experiments on the absorption of Hertzian waves by ionized gases: H. Dänzer, *Ann. d. Phys.* (5), **2**, 27–62 (1929); W. Hasselbeck, *ibid.* (5), **12**, 477–502 (1932).

Experiments on the reaction of electron-streams projected through very rarefied gas to a high-frequency field: L. Bergmann & W. Düring, *Ann. d. Phys.* (5), **1**. 1041–1068 (1929); S. Benner, *ibid.* (5), **3**, 993–996 (1929) (theory); V. Jonescu & C. Mihul, *C. R.*, **191**, 1436–1438 (1930).

Equation (24) for the so-called "plasma-electron" natural frequency appears to have been first derived by J. J. Thomson, and shortly afterward by L. Tonks & I. Langmuir, *Phys. Rev.* (2), **33**, 195–210, 990 (1929); see also Thomson, *Phil. Mag.* (7), **11**, 697–735 (1931). Experiments on this, and natural frequencies of possibly different origin: L. Tonks, *Phys. Rev.* (2), **37**, 1458–1483 (1931); H. Gutton, *Annales de physique* (10), **13**, 62–129 (1930); C. Gutton, *ibid.* (10), **14**, 5–14 (1930); Appleton & Chapman, *l.c. supra.*

Natural frequency or frequencies appearing when a constant magnetic field is imposed upon the gas: S. Benner, *Naturwiss.* **17**, 120–121 (1929); Appleton & Chapman, H. Gutton, L. Tonks, *ll.cc. supra.* The influence, upon the transmission of radio waves, of the magnetic field of the earth, was inferred independently in America by H. W. Nichols & J. C. Schelleng, *Bell Syst. Tech. Jl.*, **4**, 215–234 (1925) and in England by E. V. Appleton & M. A. F. Barnett, *Electrician*, **94**, 398 (1925).

Transformer Coupling Circuits for High-Frequency Amplifiers

By A. J. CHRISTOPHER

This article deals with the use of transformer type of coupling circuits in high-frequency amplifiers to transmit efficiently voltages or currents between certain limiting frequencies while attenuating those above and below the limiting frequencies. The similarity of these coupling circuits to band-pass filters is shown and the conditions to be satisfied in order that they may act as such are covered. Means of obtaining uniformly high amplification over relatively wide frequency bands are explained. Typical conditions under which these coupling circuits have been employed and factors affecting their performance are discussed.

I. Introduction

THE designer of high-frequency amplifiers is often confronted with the problem of obtaining, with a given number of amplifying tubes and coupling circuits in cascade, maximum voltage amplification over a predetermined band of frequencies, and high attenuation to all voltages outside of the desired band of frequencies. To secure a large voltage amplification the most convenient, economical and practical arrangement for coupling the various stages of the amplifier is by means of the step-up transformer. By adding condensers in parallel with one or more of the windings of the transformer a frequency discrimination characteristic is obtained which can be controlled to a large extent by the proper choice of the transformer constants and the tuning capacitances. In the case of the usual type of transformer coupling with the secondary winding tuned to resonance at a given frequency the voltage amplification for a single stage depends on the resistance of the secondary winding, the conductance of the grid circuit of the second tube and the size of the tuning condenser. With proper choice of the transformer constants very large amplification can generally be obtained over a relatively narrow band of frequencies. To obtain high amplifications over wider frequency bands other factors must be taken into consideration.

Possibly the most important of these factors is the impedance of the circuit into which the secondary winding operates. In the case of an interstage transformer with the secondary winding connected directly to the grid circuit of a three-element tube this impedance depends on the electrode capacities of this tube, the amplification factor, the plate impedance and the impedance connected to the output terminals of the tube. Where the amplification factor and plate circuit terminating impedance are low, as in the case of tubes operating as demodulators, the effective input impedance of the tube is comparatively high and depends mainly on the electrode capacities. The input impedance of a shielded-grid tube is likewise comparatively high and depends mainly on the capacity between the grid and filament terminals. The input impedance of a tube decreases with increase in frequency and at very high frequencies it becomes so low that difficulty is experienced in obtaining any increase in amplification in the coupling circuit. In the case of the single tuned transformer, previously referred to, the effective input capacity of the tube may act as part of the tuning condenser with the result that at the resonance frequency a substantial amplification can usually be obtained. However, when it is desired to transmit a broad band of frequencies the magnitude of the effective input capacity becomes a very important controlling factor. For a definite band width the possible voltage amplification varies inversely as the value of the effective capacity across the secondary winding.

A study of the various factors entering into the use of a suitable coupling circuit for successive stages of a high frequency amplifier has indicated that best results are obtained if the design of the transformer is based upon the principles of the broad-band filter. Definite relations are obtained between the constants of the transformer windings, the tuning capacities, and the impedances of the circuits between which it operates. The structure is thus essentially a band-pass filter and has all of its elements properly proportioned to provide the desired band selectivity, but it still retains the form and the functions of a transformer.

II. Theory and Method

Fig. 1 shows a simple transformer type of coupling circuit connected between the plate circuit of an amplifying tube and the input circuit of a second tube. The primary and secondary circuits of the trans-

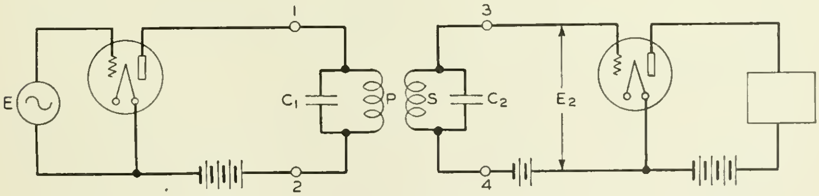


Fig. 1—Circuit schematic of an amplifier using a transformer type of coupling circuit.

former are adjusted to resonate at the same frequency. The determination of the constants for the circuit so that it conforms not only to the requirements relating to the band selectivity and to the voltage transformation ratio, but also to requirements of transmission efficiency, is explained in the following mathematical analysis.

For convenience in analyzing the essential parts of Fig. 1 they are shown in a simplified schematic form in Fig. 2. In this schematic, the sending-end impedance Z_0 corresponds to the effective plate impedance of the first vacuum tube and the electromotive force E_1 corresponds to the internal plate voltage of this tube. The transformer formed by inductances P and S is replaced by the well known "*T*" network.[1] The capacity C_1 is provided by a separate condenser, while that of C_2 is equal to the effective distributed capacity of winding S plus the effective input capacity of the second vacuum tube.

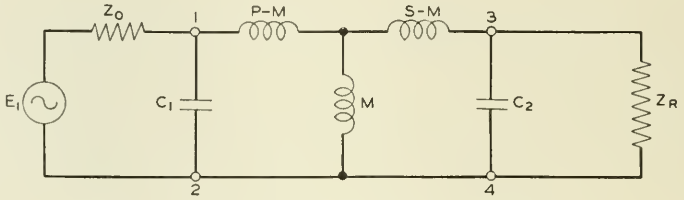


Fig. 2—"T" network of a transformer coupling circuit.

The method used in computing the output voltage of a transmission circuit such as is shown in Fig. 2 consists in determing the "image" impedances corresponding to each pair of terminals. These image impedances are determined from the open and short-circuit impedances, measured at the terminals of the network,[2] by the following relationships:

$$Z_{1,2} = \sqrt{Z_0' Z_S'}, \tag{1}$$

and

$$Z_{3,4} = \sqrt{Z_0'' Z_S''}, \tag{2}$$

where Z_0' and Z_S' equal respectively the open and short-circuit impedances and $Z_{1,2}$ the image impedance at terminals 1, 2 and Z_0'' and Z_S'' equal the open and short-circuit impedances and $Z_{3,4}$ the image impedance at terminals 3, 4.

Since the resistance elements of a transformer of this type can be made very small in comparison to the reactance elements, it is practical to eliminate them in computing the open and short-circuit impedances.

---

[1] "Telephone Transformers," by W. L. Casper, *A. I. E. E. Jour.*, March, 1924, Vol. XLIII, No. 3.

[2] "Transmission Characteristics of Electrical Wave Filters," by O. J. Zobel, *Bell Sys. Tech. Jour.*, October, 1924, Vol. III, No. 4.

With the assumption that the elements of the coupling circuit are pure reactances and that the two anti-resonant circuits are at resonance at the same frequency, f_0, the following expression for $Z_{1,\,2}$ as a function of frequency f was derived with the use of equations 1 and 2.

$$Z_{1,\,2} = \frac{1}{2\pi f C_1}\sqrt{\frac{1 - K^2}{\left(1 - K - \frac{f_0^2}{f^2}\right)\left(\frac{f_0^2}{f^2} - 1 - K\right)}} \tag{3}$$

and $Z_{3,\,4}$ is the same expression with C_2 substituted for C_1 or

$$Z_{3,\,4} = Z_{1,\,2} \times \frac{C_1}{C_2} = Z_{1,\,2}\frac{S}{P},$$

where K is the coefficient of coupling between inductances P and S, and f_0 is the common resonance frequency of the two anti-resonant circuits in Fig. 1.

The lower and upper cut-off frequencies f_1 and f_2 are related to the resonance frequency f_0 as follows:

$$\begin{aligned} f_0 &= f_1\sqrt{1 + K} \\ &= f_2\sqrt{1 - K}. \end{aligned} \tag{4}$$

The geometrical mean frequency

$$\begin{aligned} f_m &= \sqrt{f_1 f_2} \\ &= \frac{f_0}{\sqrt[4]{1 - K^2}}. \end{aligned}$$

Substituting the above expressions for f_m in place of f in equation (3), we find that at the geometrical mean frequency the image impedance in terms of f_1 and f_2 has the value

$$\frac{1}{2\pi(f_2 - f_1)C_1}, \tag{5}$$

which will be denoted by $Z_{1,\,2}$.

It will be noted upon examining equation (3) that $Z_{1,\,2}$ and $Z_{3,\,4}$ are resistive between f_1 and f_2 and are minimum at the geometrical mean frequency. They increase uniformly on either side of the geometrical mean frequency and are infinite at f_1 and f_2. For frequencies below f_1 and above f_2 they are reactive. It can therefore be seen that this type of coupling circuit inherently possesses the characteristics of a band-pass filter. However, in order that the band characteristic may be properly developed, the magnitudes of the various elements must be proportioned with respect to the terminal impedances as well as to the limiting frequencies. This is done by making the image impedances about equal to the respective terminal impedances at the mean frequency f_m of the band. Preferably the impedances should be matched at both ends, but if the transformer normally operates with one end substantially open-circuited, as in an amplifier, it is sufficient to effect the matching at the other end. The following formulas giving the constants of the various elements in terms of the limiting frequencies were derived from the foregoing equations.

$$K = \frac{f_2^2 - f_1^2}{f_2^2 + f_1^2}, \tag{6}$$

$$f_0 = \frac{\sqrt{2} f_1 f_2}{\sqrt{f_1^2 + f_2^2}}, \tag{7}$$

$$C_1 = \frac{1}{2\pi(f_2 - f_1)\, Z_{1,\,2}}, \tag{8}$$

$$P = \frac{1}{4\pi^2 f_0^2 C_1}, \tag{9}$$

$$S = \frac{1}{4\pi^2 f_0^2 C_2}, \tag{10}$$

or

$$S = P \times \frac{C_1}{C_2}\,. \tag{11}$$

As previously mentioned the above equations are based on the ideal condition which assumes the elements to be pure reactances. Actually, a small amount of resistance is present in each element of the transformer which tends to reduce the width of the transmission band. Consequently f_1 and f_2 should be assumed slightly lower and higher respectively than the lower and upper frequencies of the desired transmitted band in order to insure uniform voltage amplification. It is advantageous to make the impedance $Z_{1,\,2}$ approximately 0.8 of the terminal impedance Z_0, in which case $Z_{1,\,2}$ and Z_0 will be equal at two frequencies near the band limits and will not be greatly mismatched at the geometrical mean frequency. This tends to improve the uniformity of transmission within the band.

In applying the above equations, C_1 is determined first from equation (8). P is then determined from equation (9). From a knowledge of the effective distributed capacity of S and the input capacity of the second tube in Fig. 1, S is determined from equation (10). The windings P and S are then arranged with respect to each other to satisfy the coupling coefficient of equation (6). To facilitate adjustment C_2 may include a small adjustable condenser to compensate for variations in winding capacities as well as in the input capacity of the vacuum tube.

The following analysis shows that maximum voltage amplification is obtained with practically uniform transmission for all frequencies within the transmitted band when the secondary circuit is terminated only in its tuning condenser.

Referring to Fig. 2 and assuming that $Z_{1,\,2} = Z_0$ and $Z_R = Z_{3,\,4}$, we find that the voltage drop across Z_R is

$$E_R = \frac{1}{2} E_1 \sqrt{\frac{Z_{3,\,4}}{Z_{1,\,2}}}$$

or

$$E_R = \frac{E\mu}{2} \sqrt{\frac{Z_{3,\,4}}{Z_{1,\,2}}} \cdot \qquad (12)$$

Now from Thevenin's Theorem,[3] assuming $Z_0 = Z_{1,\,2}$, we can obtain the following expression for the output current for any value of the impedance Z_R:

$$I_r = \frac{E_2}{Z_R + Z_{3,\,4}}, \qquad (13)$$

where E_2 is the open-circuit voltage at terminal 3, 4 and I_r is the current flowing in the terminating impedance Z_R.

The actual voltage across the terminating impedance Z_R is

$$E_R = I_R Z_R = \frac{E_2 Z_R}{Z_R + Z_{3,\,4}} \cdot$$

Where Z_R is equal to $Z_{3,\,4}$ the output voltage has the value

$$E_R = \frac{E_2}{2} \cdot$$

The effect of matching the terminating impedance to the impedance $Z_{3,\,4}$ is thus to cut the output voltage in half.

Then

$$\frac{E_2}{E} = \mu \sqrt{\frac{Z_{3,\,4}}{Z_{1,\,2}}} = \mu \sqrt{\frac{C_1}{C_2}}, \qquad (14)$$

where μ = voltage amplification constant of the first tube.

Generally the effect of leaving the output end of a filter open-circuited would be to introduce irregularities of transmission within the band. However, if the band is relatively narrow the resistances of the filter elements are sufficient to smooth out these irregularities and in certain cases the transmission may be made more uniform by the omission of the impedance Z_R. This has been found to be the case with transformers of the type described here.

---

[3] "Transmission Circuits for Telephone Communication," by K. S. Johnson, p. 79.

Equation (14) shows that the voltage ratio E_2/E which is the ratio of the open-circuit voltage of the tuned circuit at terminals 3, 4 to the applied grid voltage of the first tube is directly proportional to $\sqrt{\frac{C_1}{C_2}}$ over the transmitted band. By proper choice of the constants of the coupling circuit so that $Z_{1,2}$ will equal Z_0 at a frequency lower and higher than the geometrical mean frequency this voltage ratio will remain practically constant for all frequencies within the band.

The voltage ratio E_2/E for any frequency ω outside the transmission band may be obtained from the following equation:

$$E_2/E = u\left[\frac{Z_1}{Z_2} \times \frac{Z_3}{Z_4} \times Z_5/Z_6\right],$$

where

$$Z_1 = \frac{1}{j\omega C_2}; \qquad Z_2 = j\omega(S - M) + Z_1; \qquad Z_3 = \frac{Z_2 j\omega M}{Z_2 + j\omega M};$$

$$Z_4 = Z_3 + j\omega(P - M); \qquad Z_5 = \frac{Z_4 \dfrac{1}{j\omega C_1}}{Z_4 + \dfrac{1}{j\omega C_1}}; \qquad Z_6 = Z_5 + Z_0.$$

P, S and M are in henrys and C_1 and C_2 in farads.

As the effective voltage amplification of the network within the transmission band is directly proportional to

$$\sqrt{\frac{C_1}{C_2}}$$

it is evident that since C_1 is fixed by equation (8), the value of C_2 must be kept to a minimum to realize maximum voltage step-up. Consequently, greater amplification will be obtained if the second tube in Fig. 1 has minimum input capacitance. The effective distributed capacitance of the winding is kept at a low value by placing the winding in narrow grooves on a spool.

In actual use the maximum amplification realized at the higher frequencies is somewhat less than that shown in equation (14) since the conductance component of the input impedance of the second tube reduces the effective voltage at this point. It is therefore of importance that the conductance component as well as the capacitance component of the input impedance of the second tube be kept to a minimum if maximum amplification is to be obtained. If the second tube in Fig. 1 is operated as a negative grid bias detector the ratio of E_2/E may be made considerably larger than if it is operated as an amplifier. The application of a large negative grid bias and the use of a by-pass condenser in the plate circuit of the tube to improve its modulation efficiency usually results in a very small input capacitance and conductance. It is therefore sometimes desirable in multistage amplifiers to place the selective circuits between the last amplifier tube and the detector tube. With the advent of the shielded-grid tube having the characteristic of low input capacitance and conductance, transformers of this type are particularly adapted to high frequency amplifiers such as used for the intermediate frequency amplifier of a superheterodyne receiver.

III. Practical Applications

Fig. 3 shows the transmission characteristic of a simple transformer coupling circuit when operating from a balanced resistance of 1000 ohms into the grid circuits of push-pull amplifying tubes. Stability was more important in this particular case than maximum voltage step-up so that a fixed condenser C_2 was added externally across the

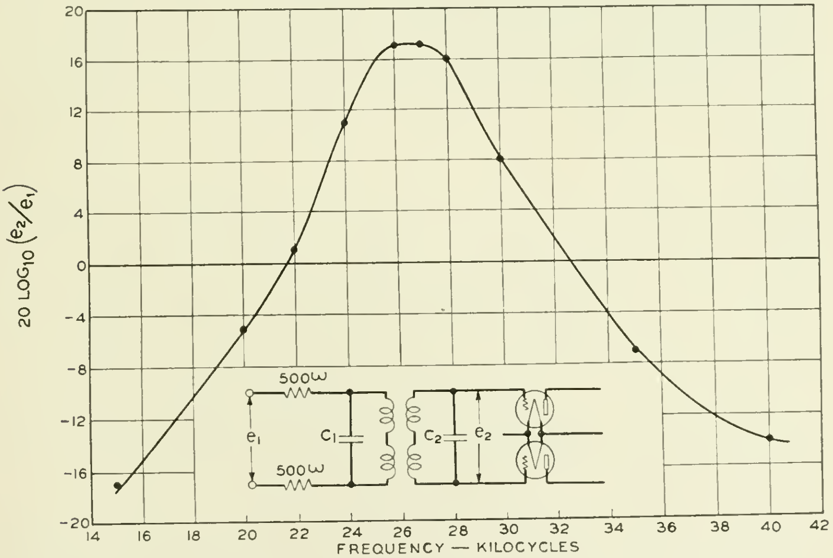


Fig. 3—Typical transmission characteristic of a transformer coupling circuit.

grid circuit to prevent the operation of the transformer from being affected by variations in the grid circuit capacities.

More than one coupling circuit has been used to obtain the necessary selectivity. Several of these have been connected together with either series or shunt condensers to obtain the equivalent of a multi-section band-pass filter. Networks of this type in which the elements have been chosen in accordance with the previously mentioned equations have been employed as band filters because of their simplicity and compactness, and the relatively low cost of the inductance elements. They have been used to connect two equal or unequal impedances as well as to operate from an impedance directly into the grid circuit of a vacuum tube.

One of these networks was used between the plate circuit of a shielded-grid vacuum tube and the grid circuit of a second shielded-grid vacuum tube at 84 kilocycles. The gain characteristic of a stage of this type is shown in Fig. 4. The capacitance C_1 represents the capacitance between the plate and shield of the first tube plus the winding capacitance and capacitance C_2 represents the winding capacitance and effective input capacitance of the second tube. It will be of interest to know that the voltage amplification obtained over the transmitted band was approximately equal to the amplification of the first shielded-grid tube. Consequently, the transformer network

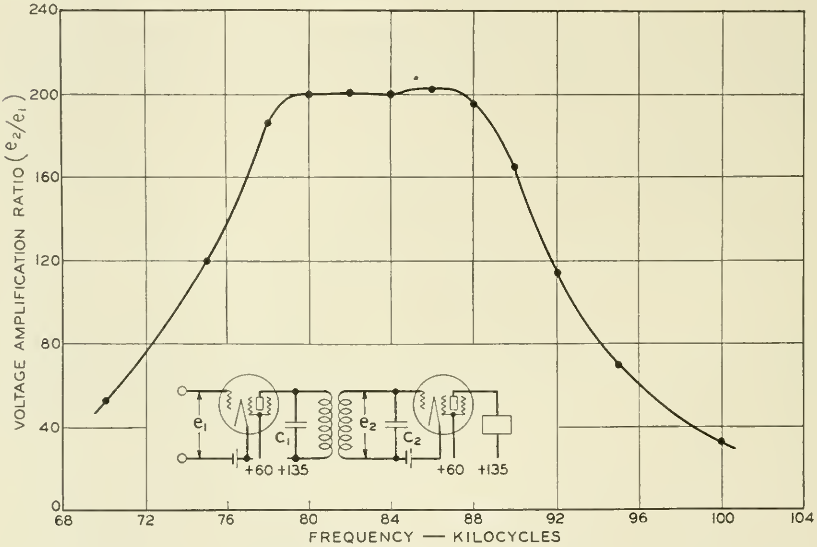


Fig. 4—Transmission characteristic of a transformer coupling circuit operating between shielded-grid tubes.

efficiently transmitted the internal plate voltage of the first tube to the grid circuit of the second tube. The internal plate impedance of the particular tube used was approximately 400,000 ohms. A two-stage amplifier consisting of two shielded-grid amplifying tubes and two transformer coupling circuits connected in cascade gave a voltage amplification of approximately 40,000 times over the frequency band shown.

Another type of transformer coupling circuit which was employed in the intermediate frequency amplifier of a high quality superheterodyne radio receiver is shown in Fig. 5. The circuit schematic of this transformer and its transmission characteristic are shown in Fig. 6. It will be noted that two transformers are mounted separately and electrically connected by a series condenser C_2'. C_1 and L_1 were determined from equations (8) and (9) and $L_1 = L_2 = L_3$. $C_2' = \frac{C_1}{2}\cdot$ The capacitance across L_4 was equal to the winding capacity and effective input capacity of the second tube. The second tube of Fig. 6 was the second detector of the intermediate frequency amplifier.

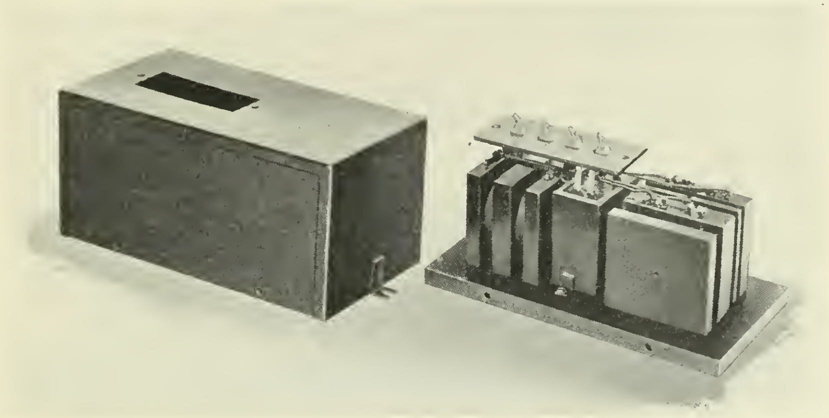


Fig. 5—Coupling circuit consisting of two-tuned transformers connected in cascade.

The elements of a coupling circuit consisting of three transformers and their associated condensers for operation over the carrier frequency range of 50 to 150 kilocycles are shown in Fig. 7. The windings of stranded wire are applied in narrow grooves to reduce the dielectric losses of the insulation between layers. The ratio of the reactance to the effective resistance for these coils varies from approximately 150 at 50 kilocycles to approximately 240 at 150 kilocycles. The halves of the winding connected to the grid circuit of the balanced tubes

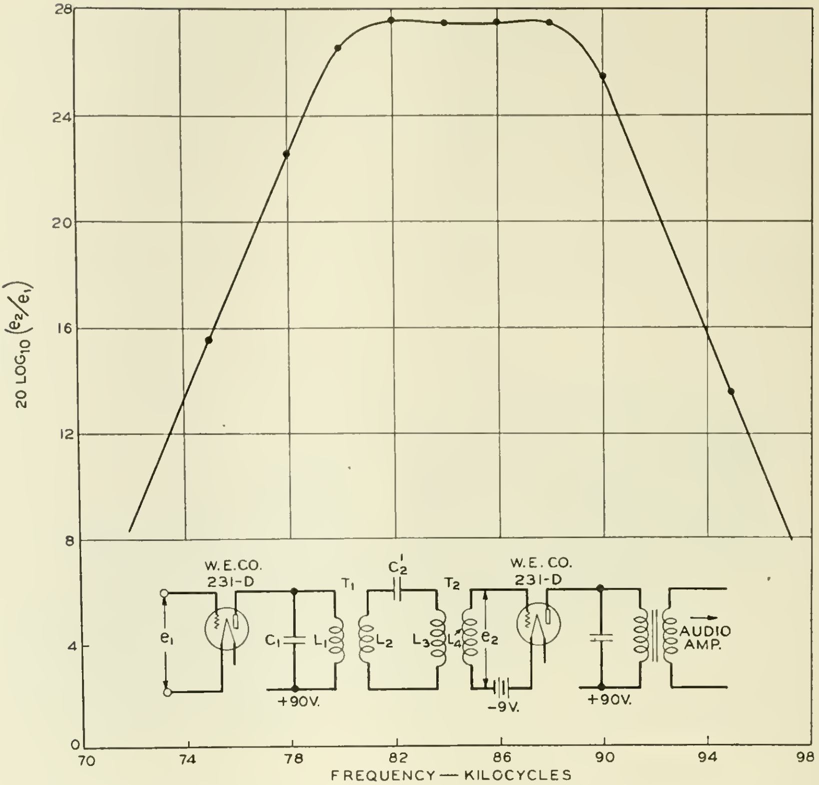


Fig. 6—Transmission characteristic of a double transformer type of tuned circuit.

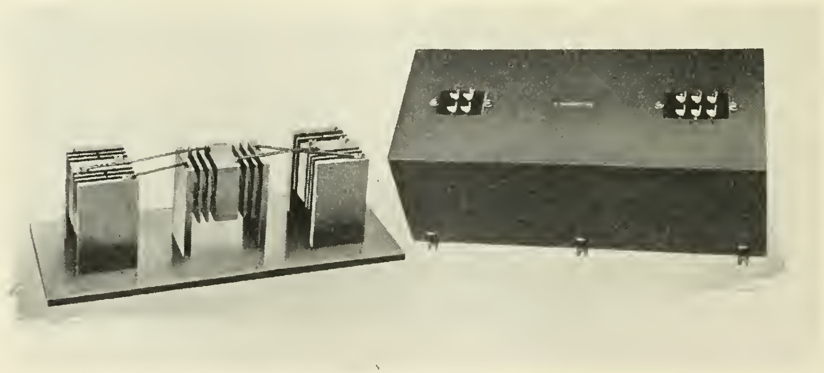


Fig. 7—Coupling circuit consisting of three-tuned transformers in cascade.

were wound in the two grooves of the last spool type assembly in such a manner as to maintain satisfactory balance for this type of operation. Four adjustable condensers were added externally to these transformers as shown in Figs. 8 and 9. The capacity of the condensers was adjustable so that the transmitted band could be located anywhere between 50 and 150 kilocycles.

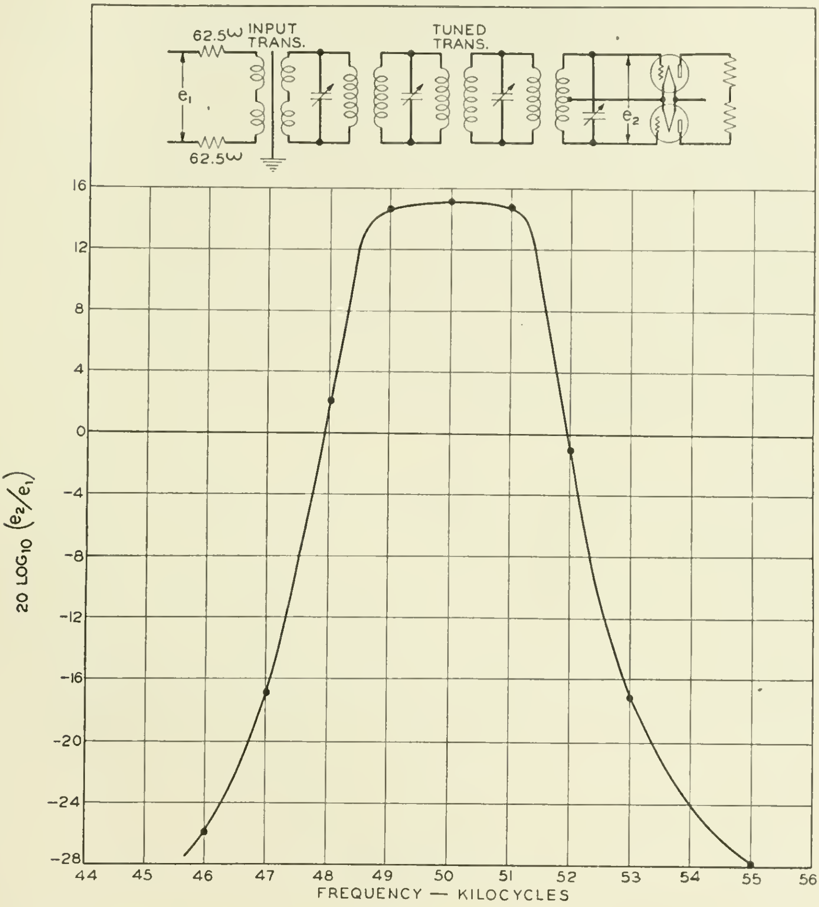


Fig. 8—Transmission characteristic of a triple transformer type of tuned circuit.

Fig. 8 shows the circuit schematic and transmission characteristic of the same transformer network at 50 kilocycles when operating from a non-inductive resistance into the grid circuit of push-pull modulator tubes. A step-up transformer was used between the resistance and the tuned transformer to match impedances and improve the efficiency of transmission. Fig. 9 shows the transmission characteristic and circuit schematic of the same tuned transformer when operating between two non-inductive resistances. Repeating coils were used between the transformer and the terminated resistances to match impedances. Although the width of the transmission band at 50

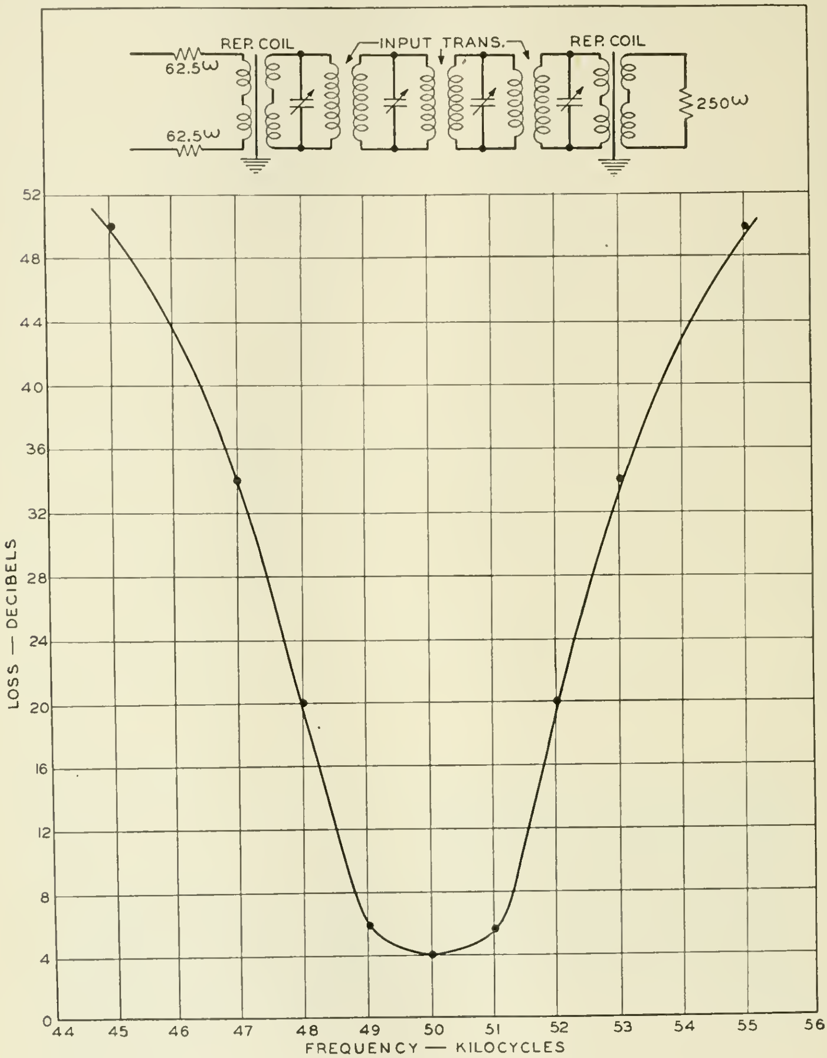


Fig. 9—Transmission loss characteristic of a triple transformer type of tuned circuit.

kilocycles is only 2500 cycles it is considerably wider at 150 kilocycles as one would expect from the equation showing the relation of band width and capacity. This, however, was not an objectionable feature in the circuit in which the transformer was employed. In order to maintain a constant band width irrespective of its location, the capacity must remain constant and the self impedance of the windings and the coupling coefficients changed or the inductance elements maintained at a constant value and the capacities and coupling coefficients changed.

IV. Conclusions

Transformer types of coupling circuits having the inductive and capacitive elements proportioned as explained in this paper are essentially band-pass filters and are therefore particularly adapted to high-frequency amplifiers or circuits where it is necessary to transmit efficiently frequencies within a desired band while strongly attenuating all frequencies outside the band. By proper choice of the transformer constants and the condensers it has been shown that uniformly high voltage amplification was obtained over relatively wide frequency bands. It has also been shown that maximum uniform voltage amplification for a given frequency band was obtained when the output terminals of a transformer coupling circuit were terminated only in a condenser.

A few applications of transformer coupling circuits have been discussed and the individual characteristic shown. It should be understood, however, that these coupling circuits are not limited to the frequency bands illustrated but may be efficiently used at higher frequencies and over wider transmission bands.

The author wishes to acknowledge the helpful suggestions of Mr. H. Whittle of the Bell Telephone Laboratories who was associated in the development of transformer coupling circuits for purposes described in this article.

Abstracts of Technical Articles from Bell System Sources

The Manufacture of Rubber-Covered Wires for Telephone Installations.[1] S. E. Brillhart. Rubber-insulated wires are extensively used by the telephone companies in connecting up apparatus and equipment which is exposed to varying climatic conditions in the same fashion in which rubber-covered wires are employed by other electrical industries. In order to meet all size, strength, and electrical requirements the various wires available for use must differ widely from one another and from commercial rubber-insulated wires, in the character of insulation with which they are covered, as well as the properties of the conductors.

At one extreme this diverse group includes wires for such use as telephone drops which extend from the cable terminals to the buildings in which the stations are located. These drop wires must be capable of carrying heavy snow and ice loads in winter and also be able to withstand exposure to summer heat and strong sunlight in hot climates. They are Nos. 14 and 17 B. & S. gauge hard-drawn copper and bronze conductors with insulations of relatively high quality containing more than 30 per cent of rubber and covered with a weatherproofed cotton braid. At the other end the group contains wires for connecting up instruments within buildings. Being supported frequently and protected from exposure, they are made from No. 22 gauge conductors, insulated with a thinner wall and covered with a colored glazed cotton braid.

A plant is located at the Point Breeze Works, Baltimore, in which certain unique methods and departures from conventional methods of manufacturing rubber-covered wires have been reduced to practice on a large scale. This paper purposes to describe the plant.

Some Recent Developments in Underground Conduit Construction in the Bell System.[2] A. L. Fox. In the past the type of joint made by trowelling cement mortar around the abutting ends of multiple clay conduit has not been entirely satisfactory because in some cases it permitted infiltration of sand and silt which obstructed the ducts.

Various types of joints have been investigated, including a modified form of the old trowelled joint, cement mortar collars molded in place, and joints made with different plastic materials such as asphalt compounds spread on a fabric backing to facilitate application and retention. None of these provided the tight seal desired.

---

[1] *Mech. Engg.*, June, 1932.

[2] *Bell Telephone Quarterly*, July, 1932.

A satisfactory joint has been developed by encasing the junction in a mortar bandage consisting of a wide band of cement mortar enclosed in cheese-cloth and held tightly against the conduit by tapes passing completely around the joint and secured on top. A separate strip of cheese-cloth is imbedded in the center of the mortar to prevent slumping. With cement mortar of proper consistency and an admixture to insure the desired plasticity sufficient cement paste comes through on the inside when the bandage is applied to provide a tight bond. A strip of paper placed in the bandage under the cheese-cloth in the outer side helps to distribute the pressure of the tying tapes and assists in retaining water in the mortar thereby aiding the hardening and increasing the strength of the joint.

The results of hydrostatic tests of these joints show that they are practically watertight. Their use is expected to effect savings in conduit construction since their high strength permits in many cases omission of the concrete base and further savings may accrue through the increased speed with which conduit can be laid and joined and the fact that the trench can be back-filled immediately without danger of injuring the joints. Other savings in the labor of rodding ducts will be realized because the new type of joint is siltproof.

The Depth of Origin of Photoelectrons.[3] Herbert E. Ives and H. B. Briggs. Previous work has shown that the photoelectrons from a silver plate covered with an equilibrium film of alkali metal follow the wave-length distribution of energy just above the silver surface, i.e., in the alkali metal. This question has been further investigated with particular references to alkali metal films in their early stages of development, where their average depth is less than one atom. Computations made on the absorption of light just within the silver surface show that there should be very definite and striking differences in the wave-length distribution of photoemission if emission occurs due to light absorption in the silver, as contrasted with emission from a film on the silver. Experimental tests made with sodium and cæsium films show that in the earliest measurable state the emission exhibits characteristics peculiar to the light absorption in silver, and that as the films build up the emission becomes characteristic of the energy above the silver. It is concluded that the photoelectrons originate partly in the underlying metal and partly in the alkali metal film, the relative proportions varying with the film thickness.

---

[3] *Phys. Rev.*, June 1, 1932.

The Lapel Microphone and Its Application to Public Address and Announcing Systems.[4] W. C. JONES and D. T. BELL. Many speakers find it difficult to use the conventional type of microphone, because of the restrictions that it imposes upon their freedom of movement. A microphone, known as the lapel microphone, designed to be attached to the speaker's clothing, has been developed for overcoming these limitations.

The vibratory structure of the lapel microphone is designed to have low mass and stiffness, and to resonate at a comparatively high frequency. The resilient support of the diaphragm adds sufficient mechanical resistance to prevent the occurrence of a prominent peak in the response at the resonance frequency. Means are provided for reducing extraneous noise to a minimum. A part of the sound reaching the microphone, due to body vibration, is rich in low frequencies and must be attenuated, otherwise the quality of transmission will be unnatural. This attenuation is accomplished in the coupling transformer, which, together with the apparatus required for suppressing clicks, for indicating when the circuit is in operation, etc., is mounted in a control cabinet. A flexible cord connects the microphone to this cabinet.

It is expected that the lapel microphone will find application in theaters, churches, convention halls, lecture and banquet rooms, and the like, where public address systems are now employed. It also can be applied in connection with other sound recording and reproducing equipment where the background noise, characteristic of carbon microphones, is not a limiting factor.

Vacuum Tube and Photoelectric Tube Developments for Sound Picture Systems.[5] M. J. KELLY. This paper reviews some recent vacuum tube and photoelectric cell developments which are of interest in sound recording and reproduction systems. An indirectly heated cathode triode is described, in the output circuit of which the current components due to the a-c. power supply of the heater have been reduced approximately 20 decibels below previously obtained levels. This tube makes it possible to use an a-c. supply in amplifiers having flat frequency characteristics with over-all gains of the order of 100 decibels. The microphonic disturbances in vacuum tubes are discussed. A measuring system for evaluating the microphonic noise currents is described, and the characteristics of a filamentary cathode tube of low microphonic noise level are given. The characteristics of a double anode, thermionic, gas-filled, rectifier tube for use in a d-c. power supply unit for the sound lamp and vacuum tube filaments of reproducing systems are given. A photoelectric cell of high sensitivity for use in sound reproduction work is described.

---

[4] *Jour. S.M.P.E.*, September, 1932.

[5] *Jour. S.M.P.E.*, June, 1932.

Analysis and Reduction of Output Disturbances Resulting from the Alternating-Current Operation of the Heaters of Indirectly Heated Cathode Triodes.[6] J. O. McNally. This paper discusses the disturbance currents in the output circuits of indirectly heated cathode triodes, introduced by the use of alternating current in the heaters. It indicates that the disturbance currents are introduced into the output circuit by (1) the electric field of the heater, (2) the magnetic field of the heater current, and (3) the resistance between heater and grid and between heater and plate, and the capacitance between heater and grid and heater and plate.

The outputs due to the electric field between cathode and plate are produced by the "grid" action of the heater and heater leads. The frequency of the output is chiefly that of the heater supply. The outputs are shown to be effectively reduced by electrostatically shielding the heater.

Disturbance currents of the frequency of the heater supply, and of double this frequency are shown to be produced by the magnetic field. The double-frequency component is shown experimentally to be proportional to the square of the heater current. The following means of reducing the magnetic field are discussed: (1) the adoption of a heater geometry which produces a smaller field in the space between the cathode and the plate, (2) the use of a magnetic shield around the heater system, and (3) the use of a lower current, higher voltage heater.

The ways in which disturbance currents are introduced by leakage resistances and capacitances between heater and grid and heater and plate are indicated, and experimental verification is given for the case of resistance between the grid and heater.

Use has been made of this disturbance current analysis in the development of an extremely low disturbance output tube, which is described.

Fourier Series in Three Dimensions.[7] W. O. Pennell. The classical Fourier Series represents a function in a given interval and then repeats the same values in the next and subsequent intervals. In other words if $S(x)$ is the classical Fourier Series representing $f(x)$ in the interval $0 < x < a$ then $S(x) = f(x - na)$ where n takes on the values $n = 0, \pm 1, \pm 2, \cdots$ corresponding to the various intervals $na < x < (n + 1)a$.

---

[6] *Proc. I.R.E.*, August, 1932.

[7] *Am. Math. Monthly*, May, 1932.

The author has shown how a generalized Fourier Series $S_1(x)$ may be obtained representing a function in the intervals $na < x < (n + 1)a$ as follows:

$$S_1(x) = b^n f(x - na),$$

where n takes on the values $n = 0, \pm 1, \pm 2, \cdots$, and b is any real constant.

In this paper is described a still more general Fourier Series $S_{11}(x)$ representing a function in the intervals $na < x < (n + 1)a$ as follows:

$$S_{11}(x) = [b^n f(x - na)]n\psi,$$

where corresponding to the above intervals n takes on the values $n = 0, \pm 1, \pm 2, \cdots$, b is any real constant, and the subponent notation $n\psi$ denotes the rotation of the plane of the curve about the X axis through an angle $n\psi$ with the XY plane.

Current Propagation in Electric Railway Propulsion Systems. JOHN RIORDAN.[8] This paper presents a systematic method of attack, based on the superposition theorem on the problems of electric railway propulsion systems arising from the presence of tracks and other leaky conductors. The treatment is limited to systems in which the tracks and other leaky conductors may be represented with sufficient accuracy by a single conductor, but includes series and shunt discontinuities in this equivalent conductor. The general equations of current propagation in a single conductor in the presence of a conductor carrying a fixed current are taken in a form similar to the transmission equations ordinarily employed for power transmission and telephone lines, apparently due to H. Pleijel, Report to Swedish Royal Railway Administration, 1919. Though these equations are not rigorous they have been found to agree with experimental observations within engineering accuracy.

The starting point of the treatment is the development of the properties of a basic circuit consisting of a straight conductor of finite length connected at its terminals to a parallel leaky conductor or track which is continuous and infinite. The circuit involves the greatest degree of continuity in the track, subject to the connection of other conductors, since the only discontinuities are those involved in the connection of the conductor to the track; it is also conveniently adapted to modifications for discontinuities. The general single series discontinuity is a fundamental point of departure in the treatment of such apparatus as track booster transformers. The basic circuit modified by discontinuities as required gives directly the propulsion circuit impedances needed in the railway network impedance diagram; it may also be employed in the construction of a particular kind of cumulative induction curve for neighboring communication lines, which takes into account the distribution of current along the track.

---

[8] Presented at the A.I.E.E. Summer Convention, Cleveland, Ohio, June 20–24, 1932. To be published in *A.I.E.E. Transactions.*

The method of building up complex propulsion systems from basic circuits is illustrated by examples chosen for their practical importance, but the paper does not give detailed procedure for engineering application of the method and its results.

Kennelly-Heaviside Layer Studies Employing a Rapid Method of Virtual-Height Determination.[9] J. P. SCHAFER and W. M. GOODALL. This paper describes a new method of determining the virtual height of the ionized regions by visual observations of the received pulse pattern on a cathode ray oscillograph tube, both for single frequencies and for two frequencies simultaneously. A résumé of the data obtained during observations of some three hundred hours is given. The frequencies used for these tests were 1604 kc, 2398 kc, 3256 kc, 4795 kc, and 6425 kc. A number of the tests included measurements made upon two frequencies in rapid rotation. The more important results may be summarized as follows:

1. On a large number of occasions during the night, a phenomenon has been observed apparently indicating an increase in the density of ionization in the lower layer. This is important because the ionization is usually assumed to decrease during the night hours.
2. Reflections are often observed simultaneously from both ionized layers. An explanation of this phenomenon is given.
3. The virtual heights of the reflecting layers are rarely duplicated from day to day for a given time and frequency.
4. Large numbers of multiple reflections are frequently obtained representing a path distance of over 5000 km. for the last reflection. This fact indicates that the multiple-hop mode of propagation is probable for long-distance transmission.

The Principles of the Light Valve.[10] T. E. SHEA, W. HERRIOTT, and W. R. GOEHNER. The light valve has been used very widely as the modulating device in systems of film sound recording. In this paper the principles of operation of the light valve are discussed, and those engineering factors which prescribe limitations on performance and indicate operating advantages are described in detail. The type of distortion which results when a light valve is overloaded is depicted both for single-plane and two-plane valves. Finally, a new type of light valve having advantages from the standpoints of weight, size, and stability of operation is described.

---

[9] *Proc. I.R.E.*, July, 1932.

[10] *Jour. S.M.P.E.*, June, 1932.

Economic Control of Quality of Manufactured Product.[11] W. A. Shewhart. This book of 501 pages is an exposition of the technique developed within the Bell Telephone System for securing economic control of quality of manufactured product at every stage in the process of fabrication all the way from raw materials to finished product. It is divided into seven parts, the first of which is devoted to a general survey of the characteristics of a controlled quality, the scientific basis for attaining control, and the economic advantages to be derived. In the second part, after a definition of what is meant by quality, the methods of presenting data are discussed in detail, both graphically and by means of statistics such as the average, standard deviation, skewness, flatness, and correlation coefficient. Part III presents the necessary and sufficient conditions for the specification of a controlled quality. The allowable sampling fluctuations in statistics are indicated in the next part, illustrated by experiments made under conditions known to be controlled. This part constitutes a survey of the present status of sampling theory. Part V takes up the problem of specifying standard quality and indicates the important changes that should be made in many kinds of specifications in order to secure the greatest assurance of uniform quality at minimum cost. Five practical criteria for determining whether or not the quality under consideration differs from standard by more than an amount that should be left to chance are presented in Part VI. In Part VII, after a summary of the fundamental principles, consideration is given to the problem of sampling, and finally a control program is presented which shows the relation of control to research, design, development, production, and purchasing. Several appendices give the original experimental data on which some of this work was based and a bibliography. Throughout the book the fundamental principles are amply illustrated by practical examples.

[11] Published by D. Van Nostrand Company, New York, 1931.

Contributors to this Issue

A. J. Christopher, B.S. in E.E., Carnegie Institute of Technology, 1921; Engineering Department, Western Electric Company, 1921–25; Bell Telephone Laboratories, 1925–. Mr. Christopher has been engaged in voice and carrier frequency transformer design and more recently concerned with the development of repeating, retardation, and induction coils.

A. B. Clark, B.E.E., University of Michigan, 1911; American Telephone and Telegraph Company, 1911–. Toll Transmission Development Engineer, 1928–. Mr. Clark's work has been largely concerned with toll telephone and telegraph systems.

Karl K. Darrow, B.S., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., University of Chicago, 1917; Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Dr. Darrow has been engaged largely in writing on various fields of physics and the allied sciences.

C. J. Davisson, B.Sc., University of Chicago, 1908; Ph.D., Princeton University, 1911; Instructor in Physics, Carnegie Institute of Technology, 1911–17; Engineering Department of the Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Dr. Davisson's work since coming with the Bell System has related largely to thermionics and electronic physics.

Bancroft Gherardi, B.Sc., Polytechnic Institute, Brooklyn, N. Y., 1891; M.E., Cornell University, 1893; M.M.E., Cornell University, 1894. New York Telephone Company, Engineering Assistant, 1895–99; Traffic Engineer, 1899–1900. New York and New Jersey Telephone Company, Chief Engineer, 1900–06. New York Telephone Company, and New York and New Jersey Telephone Company, Assistant Chief Engineer, 1906–07. American Telephone and Telegraph Company, Equipment Engineer, 1907–09; Engineer of Plant, 1909–18; Acting Chief Engineer, 1918–19; Chief Engineer, 1919–20; Vice President and Chief Engineer, 1920–. Mr. Gherardi is a Past President of the American Institute of Electrical Engineers and is now President of the American Standards Association.

Frank B. Jewett, A.B., California Institute of Technology, 1898; Ph.D., University of Chicago, 1902. American Telephone and Telegraph Company, Transmission and Protection Engineer, 1904–12. Western Electric Company, Assistant Chief Engineer, 1912–16; Chief Engineer, 1916–21; Vice President and Chief Engineer, 1921–22; Vice President, 1922–25. International Western Electric Company, Vice President, 1922–25. American Telephone and Telegraph Company, Vice President, and Bell Telephone Laboratories, President, 1925–. Dr. Jewett is a Past President of the American Institute of Electrical Engineers.

Edward C. Molina, Engineering Department of the American Telephone and Telegraph Company, 1901–19, as engineering assistant; transferred to the Circuits Design Department to work on machine switching systems, 1905; Department of Development and Research, 1919–. As Switching Theory Engineer, Mr. Molina has made contributions to the theory of probability and its applications to telephone problems, such as the efficiency of various trunking arrangements and the significance of data derived from samples.

H. S. Osborne, B.S., Massachusetts Institute of Technology, 1908; Austin Research Fellow in Engineering, 1908–10; Eng.D., 1910. American Telephone and Telegraph Company, Engineering Department, 1910–19; Department of Development and Research, 1919–20; Department of Operation and Engineering, 1920–. Mr. Osborne is Transmission Engineer and as such is responsible for assisting the Associated Companies in connection with telephone and telegraph transmission and protection matters.

The
Electrical Discoveries of
JOSEPH HENRY

By

H. S. OSBORNE AND A. M. DOWLING
AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK, U. S. A.

A Supplement to
The Bell System Technical Journal
July, 1932

The Electrical Discoveries of Joseph Henry

By H. S. OSBORNE and A. M. DOWLING

AN historical research into the accomplishments of a great scientist, such as Henry, includes two main parts. There is first the proof, based upon the record, that he accomplished certain things. This being an affirmative proof can be quite definite. There is a second step, however, which is of interest in the appraisal of his work, namely, showing in respect to what items he was first. This means in effect proving that in respect to these items no one anticipated him. Such an assertion, being a universal negative, cannot be absolutely proved, but can, of course, by sufficient work be shown to be probable in a high degree.

The student of Joseph Henry's work is hampered in the number of items which can be included in the affirmative proof by the fact that no adequate record has been preserved. Henry was slow and fragmentary in publication and most of his personal records were destroyed in a fire at the Smithsonian Institution in 1865. While his publications and the testimony of others who saw his experiments show a prodigious accomplishment, it is probable that a more complete record would have added other important contributions made by Henry to electrical science, and in other fields of scientific work.

When one attempts to appraise Henry's work in comparison with that of his contempories, and particularly to determine in what cases Henry's discoveries were prior to those of others, the difficulty is multiplied. There was a complete lack of means for rapid communication one hundred years ago. As a result scattered scientists working in the same field often independently made similar discoveries and the scientific journals in which their results were published were neither rapidly nor widely distributed. The result is illustrated by the fact that so fundamental an advance as Ohm's law, published in full in 1827, was little known or accepted even in Europe for a decade.

The original notes of the scientific investigators are for the most part, with the outstanding exception of Michael Faraday, not available, and credit for priority of discovery has generally, by common consent, been associated with priority of publication. Even at the time there were some interesting cases in Europe where conflicting claims of priority were made and the implication is made in some cases that the scientific accuracy even of the scientists themselves is open to question where matters of priority are concerned. As stated by Arago,

> "Those who discover a new fact in the sciences of observation must expect, first, to have its correctness denied,—next its importance and utility contested,—and afterwards will come the chapter of priority,—then, passages, obscure, insignificant, and previously unnoticed, will be brought forward in crowds as affording evident proofs of the discovery not being new."[1]

One hundred years later the determination of questions of priority is still difficult and it is small wonder the different writers on the subject of Joseph Henry's electrical discoveries have not always been in complete agreement as to the items for which priority should be assigned to Henry. In an attempt to resolve these uncertainties a study has been made of the published accounts of the more important discoveries of Joseph Henry in the electrical field and of the work of his contemporaries in that field. Reference has been made to the original of the publications cited by earlier investigators to indicate priority or lack of priority for Henry's work.

The results of this study are given below. It is naturally not a complete review of this complicated situation, relating as it does only to selected items of Henry's electrical work and, furthermore, other references may be found which affect the conclusions reached. It is believed, however, that this study is a contribution towards a more exact appraisal of Joseph Henry's work, and if it is of assistance to others who may in the future wish to carry this review still further, it will doubly have served a useful purpose.

In the selection of material for this review it is the intention to include in it the discoveries of Joseph Henry which are the more important from the standpoint of subsequent practical applications of electricity. The following summary indicates briefly what these discoveries were, and the conclusions in each case resulting from this study. The somewhat voluminous reference material on which the conclusions are based is given in the appendix to this paper, arranged under the same numbered headings as the summary.

Summary of Henry's Discoveries Included in This Study

1. *Construction of Electromagnets* (*1829 and 1830*)

Henry used insulated wire for the winding of electromagnets instead of the single spiral of non-insulated wire which had previously been used, and devised windings with more than one layer. By this means he was able to increase very greatly the strength of electromagnets and to make them practicable for general use. His form of construction is universal for electromagnets today. Insulated wire had previously been applied to other uses, including galvanometer coils, but not to magnets.

---

[1] *Meteorological Essays* by Francois Arago, London, 1855, p. 291.

2. Proportioning of Magnet and Battery for Maximum Effect (1830)

Henry pointed out that the coils of the magnets constructed by his method could be made up either of one very long wire or several shorter ones, showed the proper relative proportioning of magnet and battery for maximum effect in these two cases and pointed out that the intensity magnet and battery were suitable for operating over line wires of considerable resistance, while the quantity magnet and battery were suitable for local circuits. This result was arrived at experimentally before Henry had any knowledge of Ohm's law, and is the first recognition of the necessity for properly proportioning the impedance of different parts of an electric circuit which has become an important quantitative requirement in the design of all forms of electric circuit, particularly in communication circuits.

3. Electromagnetic Motor (1831)

Henry was the first to publish a description of an electric motor using electromagnets and a commutator. This motor was a reciprocating device, therefore quite different in form from the commercial motors developed later. It did, however, contribute these two fundamental elements of permanent value and also served to rouse interest in and stimulate work on the problem of obtaining mechanical power from electricity.

The question has been raised whether Henry was not anticipated by Salvatore dal Negro of Padua, who made a similar motor apparently in the same year. While the evidence on this point is somewhat circumstantial, it seems to indicate clearly that Henry was first in publication. It is probable that dal Negro's work was done without knowledge of Henry's, but there is no evidence to indicate that his motor was constructed earlier than Henry's.

4. Telegraphy (1831 and 1832)

By the application of his principles of proper proportioning of magnet winding, battery and external circuit resistance, Henry produced the first telegraph using electromagnets as a receiving device. The idea of using galvanic current for telegraphy at a distance had previously been investigated by Barlow and later the control of electromagnets at a distance by transmitting current over a line wire was investigated by Wheatstone without knowledge of Henry's work done five years earlier. In both cases the project was rejected as impracticable because of their lack of appreciation of this principle of proper proportioning.

Henry's telegraph included an "intensity" battery, a line wire of more than one mile in length and a receiving device consisting of an electromagnet with polarized armature so arranged as to strike a bell when reversals of current were sent. The use of the electromagnet made rapid signaling possible and has formed the basis of practically all subsequent commercial telegraph development. He was also the first to employ audible reception, which has played a very important part, particularly in American telegraph practice.

In Henry's later telegraph experiments he used a grounded circuit which has played a very important part in telegraph development. Grounded circuits had, however, been used before Henry in experiments with electric telegraphs of other types.

In 1850 in a Morse patent suit a claim was made that in 1830 Abraham Booth from Dublin had demonstrated a similar electromagnetic telegraph. A search of important contemporary publications, however, reveals no evidence to support this claim.

5. Electromagnetic Induction (1832)

Faraday and Henry both independently discovered the phenomena of mutual induction and of self induction. On the basis of date of publication the discovery of mutual induction is properly credited to Faraday (1831, published 1832) and the discovery of self induction to Henry (1832). The determination of this question is recognized in the official adoption by the International Electrical Congress of 1893 of the name "henry" for the practical unit of inductance.

It is interesting to note that in 1804 Vassali-Eandi observed that the shock obtained in opening a long circuit was greater than that received by merely touching the battery terminals. While this fact was due to the self inductance of a circuit this observation did not constitute an anticipation of Henry, as no thought of induction, either self or mutual, was in the mind of the investigator. Also, in connection with his investigations of mutual induction in 1831, Faraday conceived that conductors might have the property of self inductance, and carried on experiments designed to detect this property. His experiments, however, showed a negative result.

6. Electromagnetic Relay (1835)

Henry first discovered and demonstrated the principle of the electromagnetic relay. The relay was used by him at the end of his telegraph line to open the local circuit of a large magnet and to demonstrate the possibility of producing a great mechanical effect at a distance. This invention is, of course, fundamental to the operation of electrical systems of all kinds: telephone, telegraph and electric power.

7. Non-Inductive Winding (1835)

Both Henry and Faraday devised non-inductive windings in connection with their experiments on self induction. Henry's publication was, however, slightly earlier than Faraday's. The non-inductive windings so described are fundamentally the same in principle as those now used for resistance coils and other apparatus where self inductance is reduced to a minimum.

It is interesting to note that before the discovery of self induction Faraday constructed for use in certain experiments on mutual induction an apparatus closely parallel to the non-inductive winding, namely, two closely associated secondary windings connected in series opposing.

8. Transformer (1838)

While both Faraday and Henry used a transformer structure in their first mutual induction experiments, it was only later (1838, published 1839) that it was discovered by Henry that by the proper proportioning of the coils the voltage could be stepped up or stepped down with this structure. This is fundamentally the basis of modern electrical transformers.

9. Electromagnetic Shielding (1838)

Henry discovered the screening effect of all good conducting substances interposed between two coils, i.e. the primary and secondary of a transformer. After publishing his results Henry learned of Faraday's results published at about the same time purporting to show that this screening effect did not exist except with magnetic materials. Henry thereupon explained the reason for this apparent discrepancy and reaffirmed his own results.

He also pointed out that this principle has an important bearing on the improvement of electrical machinery, and explained the advantage which had been found by using bundles of wires instead of solid pieces of iron, i.e. lamination.

10. Discoveries Bearing Particularly on Radio Telegraphy and Telephony (1838 to 1842)

Henry (1838, published 1839) showed the variation in current induced in a secondary circuit as the separation between primary and secondary is varied. This is the principle of the variometer widely used in radio work.

Henry then (1842) carried his investigations to considerable distances, producing high frequency oscillating currents in a primary circuit by the discharge of a battery of Leyden jars, and detecting the result in secondary circuits at a distance first of 30 feet and then of several hundred feet. He concluded that every spark of electricity in motion exerts inductive effects at distances indefinitely great and likened the propagation of this effect to the propagation of light which he surmised was merely another kind of wave in the same medium. It is an extraordinary fact that in these experiments Henry had what would now be considered a form of radio transmitter and receiver, and that he believed the results to be produced by an electromagnetic radiation comparable with light. This was 25 years before the work of Maxwell and more than 40 years before the work of Hertz.

Henry is sometimes credited with having discovered the oscillatory nature of the discharge of the Leyden jar (1842). Felix Savary, however, in 1827 had advanced the hypothesis that this discharge is oscillatory in order to explain his experimental results. Henry repeated Savary's investigations and brought about the acceptance of Savary's hypothesis as a result of his authoritative affirmation that this hypothesis was correct.

The above record of the achievements of Joseph Henry, partial though it be, is the more impressive in consideration of the difficulties under which Joseph Henry labored. He worked in a frontier country far from the current of European scientific thought and with only occasional information regarding the results achieved by others. Furthermore, in his early years his work was limited largely to summer vacations, his time being fully occupied by teaching during the school year. His financial limitations were serious. His results were achieved by great industry in the application of his genius for practical experimentation, his inventive insight, his power for exact observation and his sense of perspective and proportion.

APPENDIX

References Bearing on Henry's Electrical Discoveries

1. *Construction of Electromagnets*

a. Use of Insulated Wire

Insulated wire was used before Henry's work for various purposes including electrostatic telegraphs and galvanometers. For example, the wire of the coil in Schweigger's galvanometer as made by Oersted was:

> ". . . wrapped or wound round throughout its whole length, in silk thread. Thus all electric communication is avoided between the different parts of this wire, the turns of which are wound one over or above the other, as it is placed in the groove of the frame."[2]

Insulated wire had not been applied to magnets, they being constructed by insulating the magnet core and winding on it a single spiral of uninsulated wire. The following is Henry's first description of the application of insulated wire to magnets:

> "In a paper, published in the *Transactions of the Albany Institute*, June, 1828, I described some modifications of apparatus, intended to supply this deficiency of Mr. Sturgeon [weak magnetic effects], by introducing the spiral coil on the principle of the galvanic multiplier of Prof. Schweigger, and this I think is applicable in every case where strong magnets cannot be used. . . . Shortly after the publication mentioned, several other applications of the coil, besides those described in that paper, were made in order to increase the size of electro-magnetic apparatus, and to diminish the necessary galvanic power. The most interesting of these, was its application to a development of magnetism in soft iron, much more extensively, than to my knowledge had been previously effected by a small galvanic element.
>
> "A round piece of iron, about 1/4 of an inch in diameter, was bent into the usual form of a horse-shoe, and instead of loosely coiling around it a few feet of wire, as is usually described, it was tightly wound with 35 feet of wire, covered with silk, so as to form about 400 turns; . . . ."[3]

The nearest approach which has been found to a contemporary statement indicating possible anticipation by others is the following published by Silliman in his journal for July, 1837:

> "Prof. Moll, of Utrecht, by winding insulated wire around soft iron, imparted to it prodigious magnetic power, so that a horse shoe bar, thus provided, and connected with a galvanic battery, would lift over one hundred pounds. About the same time, Mr. Joseph Henry, of Albany, now Prof. Henry, of Princeton College, by a new method of winding the wire, obtained an almost incredible magnetic force. . . ."

However, reference to Moll's own account of his work published in the *Edinburgh Journal of Science* April-October, 1830, and copied in *Silliman's Journal* for January, 1831, and to another contemporary account of Moll's work gives no indication of his having used insulated wire.

Henry also used other insulating materials. In SW, Vol. 1, p. 48 (1831), he acknowledged Dr. Beck's suggestion for using well waxed cotton instead of silk thread and also mentioned Dr. Beck's experiments with iron bonnet wire "which, being found in commerce already wound, might possibly be substituted in place of copper. . . ."

In 1832 he described winding the armature of a magnet with "A piece of copper wire, about thirty feet long and covered with elastic varnish. . . ."[4] In this, however, he was anticipated as to publication at least by Professor Hare of the University of Pennsylvania, who in a letter of February 24, 1831, published in *Silliman's Journal* of July 1831, stated with reference to wire for magnets, "I used no wrapping, but merely shell lac varnish, applied in winding. . . ." The same issue of *Silliman's Journal* quotes a letter from Professor J. W. Webster of Harvard of February 7, 1831, stating, "In constructing the magnetic apparatus, there is considerable economy in using sealing wax instead of silk."

---

[2] *Electro-Magnetism etc.*, Jacob Green, 1827, pp. 81–83.

[3] SW,* Vol. 1, pp. 37–38; paper published January, 1831.

[4] SW, Vol. 1 p. 75.

* *Scientific Writings of Joseph Henry*, published by the Smithsonian Institution in 1886.

b. *Multiple Layer Windings*

In 1831 Henry wrote as follows:

> "At the same time [1829 or 1830], a very material improvement in the formation of the coil suggested itself to me, on reading a more detailed account of Prof. Schweigger's galvanometer, and which was also tested with complete success upon the same horse-shoe; it consisted in using several strands of wire, each covered with silk, instead of one:—agreeably to this construction, a second wire, of the same length as the first, was wound over it, and the ends soldered to the zinc and copper in such a manner that the galvanic current might circulate in the same direction in both. . . ." [5]

This is generally acknowledged to be the first publication of the use of this important improvement in magnet construction.

In Henry's early work the windings were wound directly on the magnet core. It is not clear whether he later wound up coils separately intended to be applied to the magnet cores after winding. This form of construction, however, was used by Professor Hare as early as 1831 in making the Henry type magnets. In a letter of March 4, 1831, he described in detail the process of winding the coil on a mandril and later slipping it over the core, apparently as an improvement in construction detail.[6]

2. *Proportioning of Magnet and Battery for Maximum Effect*

Henry, in a paper published January, 1831, described a series of experiments instituted jointly with Dr. Philip Ten Eyck for the purpose of determining how the type of magnet and battery and the length of the interconnecting wires affected the magnetic force produced. From the results he concluded that:

> "From these experiments, it is evident that in forming the coil we may either use one very long wire or several shorter ones as the circumstances may require: in the first case, our galvanic combinations must consist of a number of plates so as to give 'projectile force'; in the second, it must be formed of a single pair." [7]

3. *Electromagnetic Motor*

In 1821 Faraday made the first electric motor, the action of which was the rotation of a current-carrying wire around the pole of a permanent magnet. This was followed by Barlow's motor in which a spur-shaped wheel, with current flowing through it radially, rotated between the poles of a permanent horse-shoe magnet. Henry, in an undated letter published in *Silliman's Journal* in July, 1831,[8] described in detail and illustrated an electro-magnetic rocker arm machine, provided with a commutator. This letter commences with the following statement:

---

[5] SW, Vol. 1, p. 39.

[6] *Silliman's Journal*, July, 1831.

[7] SW, Vol. 1, p. 42.

> "Sir:—I have lately succeeded in producing motion in a little machine by a power, which, I believe, has never before been applied in mechanics—by magnetic attraction and repulsion.
>
> "Not much importance, however, is attached to the invention, since the article, in its present state, can only be considered a philosophical toy; although, in the progress of discovery and invention, it is not impossible that the same principle, or some modification of it on a more extended scale, may hereafter be applied to some useful purpose. But without reference to its practical utility, and only viewed as a new effect produced by one of the most mysterious agents of nature, you will not, perhaps, think the following account of it unworthy of a place in the *Journal of Science*."

The question is sometimes raised whether Henry's motor was not anticipated by a similar one made by Salvatore dal Negro of Padua. A review of the original publications of dal Negro, while indicating that he independently invented a motor similar to Henry's, seems to indicate clearly that his dates of publication were later. The facts are these:

The first brief description of dal Negro's motor is given in a memoir entitled, "New Electro-magnetic Experiments and Observations," published in *Nuovi Saggi della Imperiale Regia Accademia di Scienze Lettere ed Arti in Padova*, Volume III, 1831. This volume is a collection of papers covering the years 1825–31, inclusive, and is one of a series of volumes issued at irregular intervals of 3 to 8 years, apparently for the purpose of providing a permanent record. This volume includes the statement that the dal Negro paper was presented before the Academy of Padua on June 21 and July 10, 1831. There is no indication how much later than this the volume was published, although the character of the volume and the other evidence in this case suggest that it may well have been published at the end of 1831 or later.

In this memoir, parts 1 to 3 inclusive, comprising 16 pages and a plate of illustrative figures, describe batteries and magnets. Part 4, only one page in length, with no figures, describes the motor very briefly as follows:

> "*Part 4—A New Electro-magnetic Motor*
>
> Having succeeded in building temporary magnets capable of holding up considerable weights by means of electromotors [batteries] of very moderate size and therefore of very small cost, I have engaged myself in the study of the way of employing this new force towards setting in motion a machine of any sort, and now I am about to

---

[8] SW, Vol. 1, p. 54.

> outline briefly to you, illustrious members of the Academy, the various manners which I have devised for putting a lever into motion.
>
> 1. The first lever I have set in motion consists of a magnetized steel bar, set vertically between the ends of a temporary magnet. The bar oscillates in response to succeeding magnetic attractions and repulsions which take place between the S pole of the bar and the S and N poles of the temporary magnet, which poles are continually reversed by the same lever. By means of the same contrivance the lever can be made to oscillate in a horizontal plane.
>
> 2. . . . ."

The rest of part 4 deals in a general way with other methods of producing motion by the use of electro-magnets and is not of interest in this discussion.

A bi-monthly publication entitled *Annali delle Scienze del Regno Lombardo—Veneto, Padova*, in its issue of July–August 1831, gives a review of the presentation of the dal Negro paper before the Academy of Padua. This review, which is more than a thousand words long, covers very completely the subject matter of the first three parts of dal Negro's paper. It makes no mention whatever of the new electro-magnetic motor. Since the description of the motor appears to be the most novel and interesting part of the paper as finally published, the omission of any mention of it in the published review naturally raises a question whether the description of the motor was a part of the paper as originally presented.

Furthermore, in the November–December 1831 issue of *Annali delle Scienze* the first three parts of dal Negro's paper are published, but not the fourth part. In this there is no reference to the electro-magnetic motor except the following statement in the introductory part, given as one of the reasons for presenting the paper:

> "4. Finally, because I hope to have enriched physics with a new motor, having succeeded in putting into motion in various manners a lever, taking advantage of the strength of temporary magnets."

The article is marked "to be continued," which suggests that it was the intention to add in a later issue a description of the electro-magnetic motor mentioned in the introduction. However, a review of subsequent issues fails to reveal any publication of part 4.

The above evidence indicates that the brief description of dal Negro's motor mentioned above probably was not published before the end of 1831, and raises considerable doubt whether it was included in the original draft of dal Negro's paper presented before the Academy in June and July, particularly in view of the lack of any mention of this part of the paper in the very full report of the meeting published in August.

The first full description of dal Negro's machine was published in the *Annali delle Scienze* for March and April 1834. This paper includes the following footnote:

> "This new instrument was made by me in the year 1831 and I have already mentioned this in the fourth part of my first memoir on temporary magnetism which I published in volume VI of the *Atti dell' Accademia di Scienze, Lettere ed Arti di Padova*." [The volume number appears to be an inaccurate reference to volume III of *Nuovi Saggi*, etc. as indicated above.]

4. *Telegraphy*

In 1831, following the description of the experiments which led to his discovery of the proper proportioning of the battery and the magnet, Henry stated:

> "But be this as it may, the fact, that the magnetic action of a current from a trough is, *at least*, not sensibly diminished by passing through a long wire, is directly applicable to Mr. Barlow's project of forming an electro-magnetic telegraph. . . ." [9]

Henry later corrected this reference to Barlow since Barlow's investigation merely tended to disprove the possibility of the telegraph. Barlow's statement regarding the telegraph was:

> "In a very early stage of electro-magnetic experiments, it had been suggested, that an instantaneous telegraph might be established by means of conducting wires and compasses. The details of this contrivance are so obvious, and the principles on which it is founded so well understood, that there was only one question which could render the result doubtful, and this was, Is there any diminution of effect by lengthening the conducting wire? . . . I was, therefore, induced to make the trial, but I found such a sensible diminution with only 200 feet of wire, as at once to convince me of the impracticability of the scheme." [10]

Even as late as 1837, Wheatstone in England had satisfied himself by experiment that the development of electro-magnetism in soft iron at a distance was impracticable. Cooke had sought to develop magnetism at a distance for the purpose of sounding an alarm. His experiments were unsatisfactory and on asking advice of Faraday and Roget was referred by the latter to Wheatstone. Wheatstone's statement was:

> "I believe, but am not quite sure, that it was on the 1st of March 1837, that Mr. Cooke introduced himself to me. . . . Mr. Cooke [later] showed me some of his drawings and models, and though I could not fully comprehend the full scope of them at the time, I saw and understood enough to assure me that his instrument was a massive complicated piece of machinery, intended to be set in action by the attractive power of an electro-magnet. On seeing this, and relying on my former experience, I at once told Mr. Cooke that it would not and could not act as a telegraph, because sufficient attractive power could not be imparted to an electro-magnet interposed in a long circuit; and to convince him of the truth of this assertion, I invited him to King's College, to see the repetition of the experiments on which my conclusion was founded. He came, and after seeing a variety of voltaic magnets, which even with powerful batteries exhibited only slight adhesive attraction, he expressed his disappointment in these words, which I well remember, 'Here is two years' labour wasted.'" [11]

Henry in 1831 or 1832 at Albany acted on his own suggestion as to

[9] SW, Vol. 1, p. 42.

[10] *Edinburgh Philosophical Journal*, Jan.–Apr. 1825.

[11] *The Electric Telegraph: Was It Invented by Professor Wheatstone?*, by W. F. Cooke, Part 2, pp. 86–87.

the applicability of his discovery to the electro-magnetic telegraph At that time he published no description of his experiments, but did so a number of years later. There are also accounts of several eyewitnesses to them, the following from a letter of James Hall to Henry, January 19, 1856.

> "While a student of the Rensselaer School, in Troy, New York, in August, 1832, I visited Albany with a friend, having a letter of introduction to you from Professor Eaton. Our principal object was to see your electro-magnetic apparatus, of which we had heard much, and at the same time the library and collections of the Albany Institute.
>
> "You showed us your laboratory in a lower story or basement of the building, and in a larger room in an upper story some electric and galvanic apparatus, with various philosophical instruments. In this room, and extending around the same, was a circuit of wire stretched along the wall, and at one termination of this, in the recess of a window, a bell was fixed, while the other extremity was connected with a galvanic apparatus.
>
> "You showed us the manner in which the bell could be made to ring by a current of electricity, transmitted through this wire, and you remarked that this method might be adopted for giving signals, by the ringing of a bell at the distance of many miles from the point of its connexion with the galvanic apparatus." [12]

And Dr. Orlando Meads, a former student of the Albany Academy, in 1863 in a discourse commemorating an anniversary of the Academy, thus referred to the scenes he had witnessed one-third of a century before:

> "The older students of the Academy in the years 1830, 1831, and 1832, and others who witnessed his experiments which at that time excited so much interest in this city, will remember the long coils of wire which ran circuit upon circuit for more than a mile in length around one of the upper rooms in the Academy, for the purpose of illustrating the fact that a galvanic current could be transmitted through its whole length so as to excite a magnet at the farther end of the line, and thus move a steel bar which struck a bell." [13]

Henry had two applications of his electro-magnets in mind: first, the production of a machine to be moved by electro-magnetism and, second, the transmission of or calling into action power at a distance.

> ". . . for the purpose of experimenting in regard to the second, I arranged around one of the upper rooms in the Albany Academy a wire of more than a mile in length, through which I was enabled to make signals by sounding a bell." [14]

Henry later used the grounded circuit, another close parallel to the form taken in the practical development of the telegraph. This principle, however, had been used years earlier in electrostatic telegraphs and in firing gun powder at a distance, e.g., by Cavallo in 1794–5 and Salva in 1798.[15]

---

[12] *Annual Report Smithsonian Institution* 1857, p. 96.

[13] *A Memorial of Joseph Henry*, published by Order of Congress, 1880, p. 380.

[14] SW, Vol. 2, p. 434, 1857.

[15] *A History of Electric Telegraphy to the Year 1837*, by J. J. Fahie, 1884, pp. 99 and 108.

In the case of Smith vs. Downing, a telegraph infringement suit tried at Boston in 1850, a witness, Oliver Byrne by name, made the following statement.

> "In the year 1830, I attended the public lectures of Abraham Booth (afterward scientific reporter for the *Times* newspaper, and who became Dr. Booth), delivered in Dublin, among other subjects, on electricity and electro-magnetism. In said lectures, the said Booth, in my presence, used in combination a long circuit of insulated wire conductors, or galvanic battery, an electro-magnet with an armature and mercury cups to join and disjoin the circuit, with which he magnetized and demagnetized the iron of the electro-magnet, causing it to attract the armature when the circuit was joined, and to recede from it when disjoined. Mr. Booth, at that time, stated to his audiences that that power could be produced and used at distant places, as signs of information; and he repeatedly illustrated what he meant, by causing the armature to approach the magnet, and then to fall from it on the floor, stating at the same time that it made marks by so falling." [16]

A search was made of various indices and publications, including *The Royal Society Catalogue of Scientific Papers*, 1800–1863; *Transactions of Royal Irish Academy*, 1786–1886; *Royal Dublin Society Catalogue*, 1731–1839, by J. F. Jones; *British Museum Catalogue;* Glazebrook's *Dictionary of Applied Physics;* J. J. Fahie's *History of the Telegraph*, previously referred to; and William B. Taylor's *An Historical Sketch of Henry's Contribution to the Electro-Magnetic Telegraph*, 1879. No further report or particulars of Abraham Booth's work, aside from the same quotation in Fahie's book, was found to support the statement and it is not therefore considered to constitute adequate evidence of anticipation of Henry.

5. *Electromagnetic Induction*

Henry's first description of his experiments on self-induction is included in his paper in *Silliman's Journal* of July, 1832. After describing experiments on mutual induction he proceeded with the following:

> "I have made several other experiments in relation to the same subject, but which more important duties will not permit me to verify in time for this paper. I may however mention one fact which I have not seen noticed in any work, and which appears to me to belong to the same class of phenomena as those before described; it is this: when a small battery is moderately excited by diluted acid, and its poles which should be terminated by cups of mercury, are connected by a copper wire not more than a foot in length, no spark is perceived when the connection is either formed or broken; but if a wire thirty or forty feet long be used instead of the short wire, though no spark will be perceptible when the connection is made, yet when it is broken by drawing one end of the wire from its cup of mercury, a vivid spark is produced. If the action of the battery be very intense, a spark will be given by the short wire; in this case it is only necessary to wait a few minutes until the action partially subsides, and until no more sparks are given from the short wire; if the long wire be now substituted a spark will again be obtained. The effect appears somewhat increased by coiling the wire into a helix; it seems also to depend in some measure on the length and thickness of the wire. I can account for these phenomena only by supposing the long wire to become charged with electricity, which by its re-action on itself projects a spark when the connection is broken." [17]

---

[16] *Historical Sketch of the Electric Telegraph*, by Alexander Jones, 1852, p. 32.

[17] SW, Vol. 1, p. 79.

Faraday's first communication on self-induction, in the *Philosophical Magazine* of November, 1834, does not state the matter very clearly. A more detailed paper read before the Royal Society on January 29, 1835, clearly presents the matter, including the following quotation:

> "The case, therefore, of the bright spark and shock on disjunction may now be stated thus: If a current be established in a wire, and another wire, forming a complete circuit, be placed parallel to the first, at the moment the current in the first is stopped it induces a current in the *same* direction in the second, the first exhibiting then but a feeble spark; but if the second wire be away, disjunction of the first wire induces a current in itself in the same direction, producing a strong spark. The strong spark in the single long wire or helix, at the moment of disjunction, is therefore the equivalent of the current which would be produced in a neighbouring wire if such second current were permitted." [18]

It is interesting to note that Faraday ascribed the momentary character of the induced current in a secondary conductor to the development in that conductor of an "electro-tonic" state, which tended to oppose the flow of current. The following interesting statement is made in a paper read on November 24, 1831:

> "The current of electricity which induces the electro-tonic state in a neighbouring wire, probably induces that state also in its own wire; for when by a current in one wire a collateral wire is made electro-tonic, the latter state is not rendered any way incompatible or interfering with a current of electricity passing through it (62). If, therefore, the current were sent through the second wire instead of the first, it does not seem probable that its inducing action upon the second would be less, but on the contrary more, because the distance between the agent and the matter acted upon would be very greatly diminished." [19]

Faraday then quoted experiments on an electrical conductor designed to detect "a return current . . . due to the discharge of its supposed electro-tonic state." These experiments were, however, unsuccessful. In view of the failure of the experiments, these speculations are not considered to constitute an anticipation of the discovery of self-induction made by both Henry and Faraday at a later date.

Another reference which is of interest regarding early observations of self-induction is the following published in the *Philosophical Magazine* in 1804 under the heading "Intelligence and Miscellaneous Articles."

> "*Velocity of the Galvanic Fluid*
>
> "Vassali-Eandi has lately made some experiments on this subject, as Beccaria did in regard to the velocity of the electric fluid. The fluid of a pile of twenty-five pairs of plates traversed in a second thirteen metres (forty-two feet and a half) of gold cord. In another experiment with a pile of fifty pairs, the fluid passed along a copper wire plated with silver, three hundred and fifty-four metres (1151 feet) in length, in a time incommensurable: the shock in this case was three times as strong as that experienced by immediately touching the two extremities of the pile." [20]

---

[18] *Experimental Researches in Electricity* by Michael Faraday, Volume 1, paragraph 1092.

[19] *Experimental Researches*, Volume 1, paragraph 74.

[20] Volume 18, p. 374.

It appears that the experimenter observed the phenomenon of momentary high voltage produced by the "extra current" later observed by Henry, but that it was in no sense an anticipation of the discovery of self-induction since it was not so interpreted, no thought of induction, self or mutual, being in the mind of the experimenter.

Henry's first publication on mutual induction was in *Silliman's Journal*, July, 1832. He mentioned having seen a short account of Faraday's work and the method he used, saying:

> "Before having any knowledge of the method given in the above account, I had succeeded in producing electrical effects in the following manner, which differs from that employed by Mr. Faraday, and which appears to me to develop some new and interesting facts."[21]

He then described his experiment, which consisted in opening and closing the circuit of an electro-magnet and noting the effect on a galvanometer connected to a coil wound around the armature of the magnet.

Faraday read a detailed paper November 24, 1831, on mutual induction, this appearing in *Philosophical Transactions*, 1832. The account Henry saw was the proceedings of the Royal Institution as reported in the *Philosophical Magazine* for April 1832, as follows:

> "Feb. 17.—Mr. Faraday gave an account of the first two parts of his recent researches in electricity; namely, volta-electric induction, and magneto-electric induction.
>
> If two wires, A and B, be placed side by side, but not in contact, and a voltaic current be passed through A, there is instantly a current produced, by induction, in B, in the opposite direction. . . . If a wire connected at both extremities with a galvanometer be coiled, in the form of a helix, around a magnet, *no current* of electricity takes place in it. . . . But if the magnet be withdrawn from or introduced into such a helix, a current of electricity is produced *whilst the magnet is in motion*, and is rendered evident by the deflection of the galvanometer. . . . Thus is obtained the result so long sought after,—the conversion of magnetism into electricity."[22]

6. *Electromagnetic Relay*

The date of Henry's first electro-magnetic relay is generally given as 1835. It was made at Princeton (where he went towards the end of 1832) after he became settled in his duties there, and before he went to England in 1837. Edward N. Dickerson,[23] a Princeton student in 1839, said that Henry put his relay in operation at Princeton in 1835. Henry's earliest description of the relay, from a deposition of 1849, is as follows:

> "In February, 1837, I went to Europe; and early in April of that year Prof. Wheatstone, of London, in the course of a visit to him in King's College, London, with Prof. Bache, now of the Coast Survey, explained to us his plans of an electro-magnetic telegraph; and, among other things, exhibited to us his method of bringing

---

[21] SW, Vol. 1, p. 75.

[22] *Philosophical Magazine*, April, 1832, pp. 300–301.

[23] *Joseph Henry and the Magnetic Telegraph: An Address by Edward N. Dickerson*, LL.D., 1885.

into action a second galvanic circuit. . . . I informed him that I had devised another method of producing effects somewhat similar. This consisted in opening the circuit of my large quantity magnet at Princeton, when loaded with many hundred pounds weight, by attracting upward a small piece of moveable wire, with a small intensity magnet, connected with a long wire circuit. When the circuit of the large battery was thus broken by an action from a distance, the weights would fall, and great mechanical effect could thus be produced, such as the ringing of church bells at a distance of a hundred miles or more, an illustration which I had previously given to my class at Princeton. . . . The object of Prof. Wheatstone, as I understood it, in bringing into action a second circuit, was to provide a remedy for the diminution of force in a long circuit. My object, in the process described by me, was to bring into operation a large quantity magnet, connected with a quantity battery in a local circuit, by means of a small intensity magnet, and an intensity battery at a distance." [24]

Under date of March 16, 1857, Henry transmitted to the Board of Regents of the Smithsonian Institution a "statement in relation to the history of the electro-magnetic telegraph" for the investigation of their special committee. Referring in his letter to his 1849 deposition he stated:

> "My testimony was given with the statement that I was not a willing witness, and that I labored under the disadvantage of not having access to my notes and papers, which were in Washington. The testimony, however, I now reaffirm to be true in every essential particular." [25]

As a result of its investigations the Committee reported as follows:

> "We have shown . . . that Mr. Henry's deposition of 1849 . . . is strictly correct in all the historical details, and that, so far as it relates to Mr. Henry's own claim as a discoverer, it is within what he might have claimed with entire justice. . . ." [26]

7. *Non-Inductive Winding*

The earliest reference to the use of the non-inductive winding by Henry was in a verbal communication made by him January 16, 1835, before the American Philosophical Society, a report on which was published March, 1835. He described it in more detail in a paper read February 6, 1835, as follows:

> "One of these ribbons was next doubled into two equal strands, and then rolled into a double spiral with the point of doubling at the centre. By this arrangement the electricity, in passing through the spiral, would move in opposite directions in each contiguous spire, and it was supposed that in this case the opposite actions which might be produced would neutralize each other. The result was in accordance with the anticipation; the double spiral gave no spark whatever, while the other ribbon coiled into a single spiral produced as before a loud snap. Lest the effect might be due to some accidental touching of the different spires, the double spiral was covered with an additional coating of silk, and also the other ribbon was coiled in the same manner; the effect with both was the same." [27]

Faraday, in his paper read before the Royal Society January 29,

---

[24] *Annual Report of Smithsonian Institution*, 1857, pp. 111–112.
[25] *Annual Report of Smithsonian Institution* 1857, p. 87.
[26] *Annual Report of Smithsonian Institution* 1857, p. 98.
[27] SW, Vol. 1, p. 94.

1835, and published in the Philosophical Transactions of that year, described the non-inductive winding as follows:

> "Thus, if a long wire be doubled, so that the current in the two halves shall have opposite actions, it ought not to give a sensible spark at the moment of disjunction: and this proved to be the case, for a wire forty feet long, covered with silk, being doubled and tied closely together to within four inches of the extremities, when used in that state, gave scarcely a perceptible spark; . . . ." [28]

Prior to this discovery Faraday in 1832 had constructed for use in certain experiments on mutual induction an apparatus closely parallel to the non-inductive winding, namely, two secondary windings twisted about each other to associate them as closely as possible and then wound in the form of a helix. With these windings connected in series opposing, he found no mutual inductive effect.[29]

8. *Transformer*

In Henry's article, "On Electro-Dynamic Induction," read November 2, 1838, and published the following year, after referring to Faraday's discovery of mutual induction he stated:

> "Since then however attention has been almost exclusively directed to one part of the subject, namely, the induction from magnetism, and the perfection of the magneto-electrical machine: and I know of no attempts, except my own, to review and extend the purely electrical part of Dr. Faraday's admirable discovery." [30]

The experiments bearing on the transformer he partially summarized as follows:

> "This experiment was considered of so much importance, that it was varied and repeated many times, but always with the same result; it therefore establishes the fact *that an 'intensity' current can induce one of 'quantity,'* and, by the preceding experiments, the converse has also been shown, that *a 'quantity' current can induce one of 'intensity.'*" [31]

9. *Electromagnetic Shielding*

Both Henry and Faraday made extensive experiments of the effect of non-magnetic shields on induction between primary and secondary coils. That the important screening effect was observed by Henry seems to have been due to his method of observation which indicated instantaneous or peak voltages, in contrast to Faraday's which gave only a measure of the total quantity of induced electric displacement.

Prior to his work on electro-magnetic shielding, Henry knew of the discovery of magnetic rotation by Arago and of the various experiments by Herschel, Babbage and Harris. These experiments had shown that a revolving magnet gives rotation to metallic discs or needles suspended above it and that this action is stopped by an interposed screen of magnetic material or thick plates of copper, silver, or zinc. Also, Faraday had shown that the rotation is due to inductive rather than magnetic action, the rotating magnet causing induced currents in the suspended body.

---

[28] *Experimental Researches in Electricity* by Michael Faraday, Vol. 1, paragraph 1096.

[29] *Experimental Researches*, Vol. 1, paragraph 194.

[30] SW, Vol. 1, p. 114.

[31] SW, Vol. 1, p. 118.

Henry did not know of the experiments on screening which Faraday had made and which were the subject of a paper read June 21, 1838, and published in the *Philosophical Transactions* for that year.[32] In this work Faraday found that a copper plate interposed between a primary and a secondary coil had no effect on the induction as indicated by a galvanometer, apparently of the slow-period type. Under this condition of measuring, the induced currents in the shields, which set up a flux opposing that of the primary coil, die out before the galvanometer response is well under way, so that the deflection is dependent only on the total change of flux in the secondary coil and hence substantially independent of the screen.

Henry, on the other hand, using shocks, which are indicative of instantaneous or peak voltages, as the measure of induction, observed the important principle of shielding, namely, that an interposed non-magnetic metallic conducting screen prevents sudden changes of flux through the screen. The "measuring instrument" in this case being instantaneous, the opposing effect of the currents induced in the shield is observed. Henry communicated these results to the American Philosophical Society November 2, 1838, a report of which, published in 1839, stated that a part of the paper:

> "relates to the effect of interposing different substances between the conductor which transmits the current from the battery, and that which is arranged to receive the induced current. All good conducting substances are found to screen the inducing action, and this screening effect is shown, by the detail of a variety of experiments, to be the result of the neutralizing action of a current induced in the interposed body." [33]

Henry, on later learning of Faraday's results and of the then apparent discrepancy between those and his own, investigated the subject further with additional experiments, and from the experimental data and theoretical reasoning arrived at a fair reconciliation of the various observations.[34]

Speaking of the phenomena of circulating currents induced in an interposed shield, Henry said:

> "Also the same principle appears to have an important bearing on the improvement of the magneto-electrical machine: since the plates of metal which sometimes form the ends of the spool containing the wire, must necessarily diminish the action,

---

[32] *Experimental Researches in Electricity* by Michael Faraday, Vol. 1, paragraphs 1709–1736.

[33] SW, Vol. 1, p. 106.

[34] SW, Vol. 1, pp. 169–188.

> and also from experiment of paragraph 72 the armature itself may circulate a closed current which will interfere with the intensity of the induction in the surrounding wire. I am inclined to believe that the increased effect observed by Sturgeon and Calland, when a bundle of wire is substituted for a solid piece of iron, is at least in part due to the interruption of these currents. I hope to resume this part of the subject, in connection with several other points, in another communication to the Society." [35]

At the end of the meeting in which this communication was presented, Dr. Bache gave an account of investigations of Professor Ettingshausen of Vienna, in which he had been led to suspect the development of a current in the metal of the keeper of the magneto-electrical machine, which diminished the effect of the current in the coil about the keeper. Mentioning this communication, Henry said:

> "It gives me pleasure to learn that the improvements, which I have merely suggested as deductions from the principles of the interference of induced currents (76), should be in accordance with the experimental conclusions of the above named philosopher." [36]

Under date of November 13, 1838, a few days after Henry's communication to the American Philosophical Society, a communication bearing on this subject was sent to *Silliman's Journal* by Charles G. Page and published in that *Journal* in January, 1839. Mr. Page states:

> ". . . I would not say that a very careful insulation [of the iron wires forming a magnet core] might not improve their operation. For I apprehend that in the development and return of magnetic forces, electrical currents are excited in the body of the magnet at right angles to its axis, as well as in the wires surrounding the magnet."

The work of Foucault, whose name is often associated with this type of induced current, was published in *Comptes Rendus*, 1855. Foucault was familiar with the work of Arago and Faraday, but did not mention the work of Henry.

10. *Discoveries Bearing Particularly on Radio Telegraphy and Telephony*

a. *Variation of Induction Between Coils with Separation*

Henry, in his extensive investigations of induction, as presented in a paper in 1838 and published in 1839,[37] showed that the current induced in the secondary of two helical coils could be varied or regulated by changing the relation in space of the coils. The following from several of Henry's applications of this principle gives a partial summary of the results of his investigations:

> "The action at a distance affords a simple method of graduating the intensity of the shock in the case of its application to medical purposes. The helix may be sus-

---

[35] SW, Vol. 1, p. 126.
[36] SW, Vol. 1, p. 145.
[37] SW, Vol. 1, pp. 119–122.

pended by a string passing over a pulley, and then gradually lowered down towards the plane of the coil, until the shocks are of the required intensity." [38]

b. Action of Induction at a Distance

Henry investigated the action of the discharge of a Leyden jar through a primary circuit in inducing currents in distant secondary circuits. For example, the following report of a paper read in 1842 was published the following year:

> ". . . a single spark . . . thrown on the end of a circuit of wire in an upper room, produced an induction sufficiently powerful to magnetize needles in a parallel circuit of wire placed in the cellar beneath, at a perpendicular distance of thirty feet with two floors and ceilings . . . intervening. . . . . . . when it is considered that the magnetism of the needle is the result of the difference of two actions, it may be further inferred that the diffusion of motion in this case is almost comparable with that of a spark from a flint and steel in the case of light." [39]

These experiments were carried to greater distances as is illustrated by the following excerpt from the proceedings of the meeting of the American Philosophical Society, October 21, 1842:

> "Prof. Henry communicated, orally, an extension of the experiments, which he had previously brought before the Society, on electro-dynamic induction. He had succeeded in magnetizing needles by the secondary current in a wire more than two hundred and twenty feet distant from the wire through which the primary current was passing, excited by a single spark from an electrical machine." [40]

Also, in the notes on "Lectures on Natural Philosophy by Professor Henry" made by William J. Gibson, a student, February 28, 1844, after a description of experiments on induction at a distance in which inductive effects were noted several hundred feet away, it is stated:

> "Hence the conclusion that every spark of electricity in motion exerts these inductive effects at distances indefinitely great (effects *apparent* at distances of one-half a mile or more); and another ground for the supposition that electricity pervades all space. Each spark sent off from the Electrical Machine in the College Hall sensibly affects the surrounding electricity through the whole village. A fact no more improbable than that light from a candle (probably merely another kind of wave or vibration of the same medium), should produce a sensible effect on the eye at the same distance." [41]

This is also referred to in Henry's letter to Rev. S. B. Dod, written in 1876 as follows:

> "As another illustration of this, it may be mentioned that when a discharge of a battery of several Leyden jars was sent through the wire before mentioned, stretched across the campus in front of Nassau Hall, an inductive effect was produced in a parallel wire, the ends of which terminated in the plates of metal in the ground in the back campus, at a distance of several hundred feet from the primary current, the building of Nassau Hall intervening. The effect produced consisted in the magnetization of steel needles." [42]

[38] SW, Vol. 1, p. 121.

[39] SW, Vol. 1, p. 203.

[40] *Proceedings of the American Philosophical Society*, Vol. 2, Jan., 1841–May, 1843.

[41] Manuscript in Princeton Library, p. 135.

[42] *A Memorial of Joseph Henry*, published by order of Congress, 1880, pp. 151–152.

The effect of lightning on an electroscope had been observed prior to Henry's work. In extending this work, Henry connected a wire to the metallic roof of his study, carried this wire through a magnetizing spiral and to ground, by means of which arrangement needles were magnetized in his study "even when the flash is at a distance of seven or eight miles, and when the thunder is scarcely audible."[43] William J. Gibson, in 1844, in his notes on Henry's lectures, states that needles were magnetized by lightning 20 miles away.

c. Oscillatory Nature of Discharge of Leyden Jar

Anomalies in the polarity of small steel needles magnetized by the discharge from the Leyden jar had been observed prior to Henry's work notably by Felix Savary, who in 1827 advanced in the following words the hypothesis that these results were due to the oscillatory discharge of the Leyden jar:

> "An electric discharge is a phenomenon of movement. Is this movement a continuous translation of matter in a determined direction? Then the opposite polarity of magnetism observed at different distances from a straight conductor, or in a helix with gradually increasing discharges, would be due entirely to the mutual reactions of the magnetic particles in the steel needles. The manner in which the action of a wire changes with its length appears to me to exclude this supposition.
>
> "Is the electric movement during the discharge, on the other hand, a series of oscillations transmitted from the wire to the surrounding medium and soon attenuated by resistances which increase rapidly with the absolute velocity of the moving particles?
>
> "All the phenomena lead to this hypothesis which makes not only the intensity but the polarity of the magnetism depend on the laws in accordance with which the small movements die out in the wire in the surrounding medium and in the substance which receives and conserves the magnetism."[44]

Henry repeated and extended Savary's investigations and confirmed Savary's hypothesis by a clear and authoritative statement of the phenomenon as follows:

> "The discharge, whatever may be its nature, is not correctly represented (employing for simplicity the theory of Franklin) by the single transfer of an imponderable fluid from one side of the jar to the other; the phenomena require us to admit *the existence of a principal discharge in one direction, and then several reflex actions backward and forward, each more feeble than the preceding, until the equilibrium is obtained.* All the facts are shown to be in accordance with this hypothesis, and a ready explanation is afforded by it of a number of phenomena which are to be found in the older works on electricity, but which have until this time remained unexplained."[45]

References

For convenience there is given below a list of the principal publications referred to in connection with the study as original sources. Many of the references cover the general field of electrical discovery contemporary with Henry; others relate to specific items of the work.

---

[43] SW, Vol. 1, p. 203—Paper presented in 1842 and published the following year.

[44] *Annales de Chimie et de Physique*, 1827.

[45] SW, Vol. 1, p. 201, 1842.

A number of these references are mentioned in the discussion of specific items. The list does not include thirty or forty present day publications relating to Henry's work, which were reviewed, from many of which valuable suggestions were obtained as to where to look for source material, but which were not themselves used as sources.

Scientific Writings of Joseph Henry, published by the Smithsonian Institution, 1886.
Experimental Researches in Electricity, by Michael Faraday, D.C.L., F.R.S., 1839.
A Memorial of Joseph Henry, published by Order of Congress, 1880.
Annual Report of the Smithsonian Institution for 1857.
Students' Notes on Henry's Princeton Lectures, MSS. in Princeton Library.
The Electric Telegraph: Was It Invented by Professor Wheatstone?, by William Fothergill Cooke, London, 1857.
Historical Sketch of the Electric Telegraph, by Alexander Jones, New York, 1852.
A History of Electric Telegraphy to the Year 1837, by J. J. Fahie, London, 1884.
An Historical Sketch of Henry's Contribution to the Electro-Magnetic Telegraph, by William B. Taylor, 1879.
Joseph Henry and the Magnetic Telegraph: An Address, by Edward N. Dickerson, LL.D., 1885.
The Annals of Electricity, Magnetism, etc., by William Sturgeon, 1836–1843, London.
Scientific Memoirs, Edited by Richard Taylor, F.S.A., London, 1837–1852.
Electro-Magnetism, etc., by Jacob Green, M.D., Philadelphia, 1827.
Meteorological Essays by Francois Arago, London, 1855.
Glazebrook's Dictionary of Applied Physics.
Annali delle Scienze del Regno Lombardo—Veneto, Padova, 1831–1834.
Nuovi Saggi della Imperiale Regia Accademia di Scienze Lettere ed Arti, in Padova, Vols. I–VI, 1817–1847.
American Journal of Science (Silliman's Journal), 1822 and 1827–1842.
Proceedings of American Philosophical Society, Vols. 1–3, and Vol. for 1744–1838.
Transactions of American Philosophical Society, Vols. VI–VIII (1834–1842).
Philosophical Transactions, London, 1831, 1832, 1834, 1835, and 1838.
Philosophical Magazine, London, 1803, 1804, 1828–1832 and 1834.
Edinburgh Philosophical Journal, 1825.
Edinburgh Journal of Science, 1830.
Annales de Chimie et de Physique, 1827.
Comptes Rendus, 1855.
Royal Society Catalogue of Scientific Papers, 1800–1863.
Transactions of Royal Irish Academy, 1786–1886.
Royal Dublin Society Catalogue of the Library, 1731–1839, by J. F. Jones.
British Museum Catalogue.